

---

# The Philosophy Major's Introduction to Philosophy

---

Concepts and Distinctions

Ken Akiba

First published 2021  
by Routledge  
52 Vanderbilt Avenue, New York, NY 10017

and by Routledge  
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

*Routledge is an imprint of the Taylor & Francis Group, an Informa business*

© 2021 Taylor & Francis

The right of Ken Akiba to be identified as author of this work has been asserted by him in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilized in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

*Trademark notice:* Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

*Library of Congress Cataloging-in-Publication Data*

Names: Akiba, Ken, author.

Title: The philosophy major's introduction to philosophy:  
concepts and distinctions / Ken Akiba.

Description: New York, NY: Routledge, 2020. |

Includes bibliographical references and index. |

Identifiers: LCCN 2020014970 (print) | LCCN 2020014971 (ebook) |

ISBN 9780367482985 (hardback) | ISBN 9780367482978 (paperback) |

ISBN 9781003039136 (ebook)

Subjects: LCSH: Philosophy--Introductions.

Classification: LCC BD21 .A35 2020 (print) |

LCC BD21 (ebook) | DDC 100--dc23

LC record available at <https://lcn.loc.gov/2020014970>

LC ebook record available at <https://lcn.loc.gov/2020014971>

ISBN: 978-0-367-48298-5 (hbk)

ISBN: 978-0-367-48297-8 (pbk)

ISBN: 978-1-003-03913-6 (ebk)

Typeset in Bembo  
by Newgen Publishing UK

---

# Contents

---

<i>List of figures</i>	x
<i>Preface</i>	xiii

1	Particulars and Universals; Logic and Language	1
1.1	<i>Tokens and Types; Particulars and Universals</i>	1
1.2	<i>Realism and Anti-realism</i>	4
1.3	<i>Propositional Logic</i>	5
1.4	<i>Predicate Logic</i>	8
1.5	<i>Identity</i>	12
1.6	<i>Necessary and Sufficient Conditions</i>	19
1.7	<i>Quotation</i>	21
	<i>Exercise Questions</i>	25
	<i>Suggested Further Reading</i>	29
	<i>Notes</i>	30
2	Extension and Intension	34
2.1	<i>Introduction</i>	34
2.2	<i>Set Theory</i>	34
2.3	<i>Mereology</i>	38
2.4	<i>Kinds of Extension and Intension</i>	42
2.5	<i>Possible Worlds</i>	47
2.6	<i>Mathematical Functions</i>	49
2.7	<i>The Possible Worlds Analysis of Intension</i>	51
2.8	<i>Rigid Designators</i>	55
2.9	<i>A Problem with the Possible Worlds Analysis of Intension</i>	57
	<i>Exercise Questions</i>	59
	<i>Discussion Question</i>	61
	<i>Suggested Further Reading</i>	61
	<i>Notes</i>	61

3	Analyticity, Apriority, and Necessity	64
3.1	<i>Four Distinctions in Truths</i>	64
3.2	<i>Logical vs Non-logical Truths</i>	65
3.3	<i>Analytic vs Synthetic Truths</i>	67
3.4	<i>A Priori vs A Posteriori Truths</i>	69
3.5	<i>The Possible Worlds Analysis of Modality; Modal Logic</i>	70
3.6	<i>Metaphysical Modality; the Necessary Truth</i>	72
3.7	<i>Essence and Haecceity</i>	73
3.8	<i>The Puzzle about the Statue and the Clay</i>	74
3.9	<i>De Re and De Dicto Modality</i>	75
3.10	<i>'The Trinity Thesis'</i>	77
3.11	<i>Kant's Synthetic A Priori</i>	78
3.12	<i>Kripke's Necessary A Posteriori</i>	82
3.13	<i>Counterfactual Conditionals</i>	85
3.14	<i>Causation</i>	88
3.15	<i>Epistemic and Deontic Modality</i>	89
3.16	<i>Temporal Modality</i>	94
	<i>Exercise Questions</i>	96
	<i>Discussion Questions</i>	98
	<i>Suggested Further Reading</i>	99
	<i>Notes</i>	100
4	Content, Linguistic and Mental	103
4.1	<i>Form and Content; Linguistic and Mental Content</i>	103
4.2	<i>Propositional Attitudes</i>	105
4.3	<i>Extensional and Intensional Contexts</i>	109
4.4	<i>De Re and De Dicto Mental Content</i>	112
4.5	<i>Descartes's Argument for Dualism</i>	113
4.6	<i>Skepticism; 'a Brain in a Vat'</i>	116
4.7	<i>Moral Error Theory</i>	118
4.8	<i>Performative Utterances</i>	120
4.9	<i>Moral Expressivism and the Frege-Geach Problem</i>	122
	<i>Exercise Questions</i>	126
	<i>Discussion Questions</i>	127
	<i>Suggested Further Reading</i>	128
	<i>Notes</i>	129
5	Internalism and Externalism	130
5.1	<i>Internalism vs Externalism in General</i>	130
	<b>Part A Semantic Internalism and Externalism</b>	<b>131</b>
5.2	<i>The Description Theory of the Reference of Proper Names</i>	131
5.3	<i>Kripke's Criticism of the Description Theory</i>	133
5.4	<i>The Causal Theory</i>	135

5.5	<i>Searle's Defense of the Description Theory</i>	136
5.6	<i>The Meaning of Natural Kind Terms; 'Twin Earth'</i>	139
5.7	<i>Two Internalist Responses</i>	140
5.8	<i>Narrow vs Wide Content; Intrinsic vs Extrinsic Properties</i>	142
5.9	<i>Supervenience</i>	144
	<b>Part B Internalism and Externalism in Epistemology</b>	<b>147</b>
5.10	<i>The JTB Theory of Knowledge</i>	147
5.11	<i>Internalist Theories: Foundationalism and Coherentism</i>	149
5.12	<i>An Externalist Theory: Reliabilism</i>	152
5.13	<i>Putnam's Semantic Externalist Argument against Skepticism</i>	156
	<b>Part C Internalist and Externalist Elements in Personal Identity</b>	<b>158</b>
5.14	<i>Locke's Theory of Personal Identity</i>	158
5.15	<i>Genuine and Pseudo Memory</i>	160
5.16	<i>The Duplication Problem and the No Competitor Theory</i>	162
	<i>Exercise Questions</i>	165
	<i>Discussion Questions</i>	165
	<i>Suggested Further Reading</i>	166
	<i>Notes</i>	167
	 <i>Answers to Selected Exercise Questions</i>	 171
	<i>Bibliography</i>	177
	<i>Index</i>	181

---

# Figures

---

1.1	Words on the whiteboard	1
1.2	Animals in my backyard	2
1.3	A four-dimensional object	16
1.4	The Ship of Theseus	17
1.5	Adam said, 'I will go to school today'	23
1.6	Finger quotes	24
2.1	$a \in S$	35
2.2	$a \notin S$	35
2.3	$\neg P$	36
2.4	$P \cap Q$	37
2.5	$P \cup Q$	37
2.6	$P \subseteq Q$	37
2.7	$P \subset Q$	38
2.8	The set of my body parts	39
2.9	The same set of my body parts	39
2.10	My body as a sum of my body parts	40
2.11	A different sum of my body parts	41
2.12	The square function	50
2.13	The 'x is a dog' function	51
2.14	Possible denotations	52
2.15	Possible extensions	52
2.16	Synonymous predicates	53
2.17	Rigid designators	56
2.18	Rigid designators co-extensional in the actual world	57
2.19	Synonymous predicates?	58
3.1	The Trinity Thesis	77
3.2	Parallel Postulate	79
3.3	A model of two-dimensional hyperbolic geometry	80
3.4	A model of two-dimensional elliptic geometry	81
3.5	Epistemic possible worlds	90
3.6	Deontic possible worlds	93
3.7	Temporal worlds	94

4.1	A brain in a vat	117
4.2	'The signal is green'	118
4.3	'Abortion is morally wrong'	119
5.1	The pyramid	149
5.2	The raft (or web)	151
5.3	'I am a brain in a vat'	156
5.4	Memory link	159

---

# Preface

---

The undergraduate philosophy curriculum at a typical American university is divided into two: the lower-level courses for general education of freshmen and sophomores and the upper-level courses for philosophy majors. By design, the former are easier and more accessible, aiming at a self-contained introduction to philosophy for a general student population who may never take another course in philosophy, while the latter are more challenging and substantive, focusing on giving majors a comprehensive and intensive study of philosophy for at least two years and possibly more. What is often the case, however, is that those introductory courses at the lower level, due to their self-contained nature, do not sufficiently prepare majors for further sustained study at the upper level. Many philosophy majors (and minors) who are transitioning from the lower level to the upper level are shocked by the gap between the relative ease of lower-level philosophy courses and the difficulty of upper-level philosophy courses. They often feel as if they haven't obtained the knowledge and skills they are expected to have for advanced study. Instructors are bogged down and frustrated in having to teach requisite elementary concepts before discussing the original sources they want to discuss. Moreover, upper-level philosophy courses are usually divided into specialized areas such as ethics, metaphysics, epistemology, philosophy of mind, philosophy of language, and philosophy of science; as a result, it is often difficult for majors to obtain a broader view of philosophy in general.

*The Philosophy Major's Introduction to Philosophy* aims to mitigate these problems by offering rigorous but concise and accessible accounts of basic philosophical concepts and distinctions which are used throughout different areas of analytic philosophy. Concepts and distinctions to be discussed include the following. (You don't need to know or understand these terms *now*. You study this book to understand them.)

- particulars/universals;
- tokens/types;
- abstract/concrete objects;



- realism/anti-realism;
- singular terms/predicates;
- numerical/relative identity;
- eternal/temporary identity;
- necessary/contingent identity;
- necessary/sufficient conditions;
- use/mention;
- object language/metalinguage;
- sets/mereological sums;
- extension/intension;
- properties/relations/propositions;
- intrinsic/extrinsic properties;
- mental/physical properties;
- essential/accidental properties;
- essence/haecceity;
- internal/external relations;
- possible worlds;
- rigid designators;
- logical/non-logical truths;
- analytic/synthetic truths;
- a priori/a posteriori knowledge and truths;
- necessary/possible/contingent truths;
- de re/de dicto modality and mental content;
- material/indicative/subjunctive (or counterfactual) conditionals;
- metaphysical/epistemic/deontic/temporal modality;
- form/content;
- syntax/semantics;
- linguistic/mental content;
- propositional attitudes;
- extensional/intensional contexts;
- dualism/physicalism (or materialism);
- skepticism;
- error theory/expressivism;
- performative utterances;
- internalism/externalism;
- description theory/causal theory;
- narrow/wide content;
- supervenience;
- foundationalism/coherentism/reliabilism;
- psychological continuity (or memory) theory of personal identity.

Brief but sufficient descriptions of such logico-mathematical tools as propositional logic, predicate logic, modal logic, set theory, mereology, and mathematical functions are incorporated. The core of this book probably

belongs to the area in which philosophy of language, metaphysics, and logic overlap, as that's the origin of many of the most important concepts in contemporary analytic philosophy, but at least one topic in ethics (moral anti-realism) and two topics in epistemology (skepticism and the internalism–externalism debate about epistemic justification) are also included. I also tried to make connections with modern philosophy and philosophers, and included discussions about Descartes's skepticism and dualism, Locke's theory of personal identity, Hume's theory of causation, and Kant's synthetic *a priori*, as I suspect that's the area the majority of targeted readers of the book are already familiar with. Some well-known entertaining puzzles and thought experiments such as the Ship of Theseus, the Statue and the Clay, a Brain in a Vat, and Twin Earth are also included.

While this book can be useful to those who wish just to look up one or another of those concepts and distinctions, it has a continuous narrative and is generally meant to be studied as a whole. The book begins with the twenty-five-century-old distinction between particulars and universals. Universals are species of abstract objects, i.e., objects that exist outside of spacetime. Chapter 1 introduces singular terms and predicates as linguistic expressions that represent particulars and universals, respectively. Chapter 2 then introduces a more subtle distinction between extensions and intensions of linguistic expressions and connects them with the previous discussion about abstract objects. The chapter defines propositions as things that exist independently of our language but that can be expressed as intensions of sentences. Possible worlds are introduced for the analysis of intension. Chapter 3 deals with four distinctions in truths (i.e., true sentences or propositions): logical vs non-logical truths, analytic vs synthetic truths, *a priori* vs *a posteriori* truths, and necessary vs contingent truths. The discussion about possible worlds and modality will continue. Chapter 4 introduces the idea of mental representation. Just as sentences express propositions, propositional attitudes grasp propositions as their content. But propositions can be true or false (or neither). That's where many philosophical issues arise, such as knowledge and skepticism, and error theory and expressivism. Finally, Chapter 5 contains three topics involving the internalism/externalism distinction: internalism vs externalism about meaning and content, about epistemic justification, and about personal identity. The first topic addresses the question about how our linguistic and mental content is determined.

This book can be used as the (or a) textbook for a one-semester college course or for self-study. In either case, the expected reader is somebody who has already taken a few general introductory courses in philosophy, such as an introduction to philosophy, an introduction to ethics, modern philosophy, and an introductory course in propositional logic. (In fact, I assume the reader's prior familiarity with propositional logic in Chapter 1 even though, in my experience, students who don't have such familiarity can easily catch up with it.) This book is not a lower-level introduction to philosophy

textbook. Not only is its content too difficult for most newcomers, but, more importantly, it does not aim to motivate them or entice them into studying philosophy; it is a book for those who are already interested in philosophy and want to study more, like philosophy majors.

An ideal course for the book would be a gateway course to the upper-undergraduate philosophy courses, a course some may call ‘pro-seminar’. (I am aware that many philosophy programs do not have such a course; but I think they should.) But, alternatively, it can be used in various crossover courses. In a one-semester course, the instructor can combine a teaching of the book with writing assignments and/or student research on some of the topics discussed in the book. A classroom reading of some pertinent readings, such as excerpts from Frege and Kripke, could also be included, but is not necessary. The book is compact, but there is much content in it, and the teaching of the whole book should leave little time to spare for any additional class activities in a one-semester course. Some of the later sections of each chapter (except Chapter 2) can be skipped at the instructor’s discretion.

If you want to use this book for self-study, you are strongly advised to read it slowly and carefully, from the beginning to the end without skipping parts (as you don’t know exactly what can be skipped). Answer Exercise Questions at the end of the chapters (and, if possible, answer Discussion Questions, too). Keep in mind that the book is no substitute for upper-undergraduate textbooks in philosophy of language, philosophy of mind, metaphysics, or epistemology, let alone in other areas; it covers much less. It aims to survey those areas in a continuous narrative; it does not aim to exhaust all the important theories in each subject. I did not even try to always present the most popular theory in the subject; instead, I tried to present the theory which is most intelligible and helpful to the readers at the targeted level. If you become interested in the subject, study more. The papers and books listed in the Suggested Further Reading sections may help.

More specifically, I omitted any discussion about (among other things) various theories of universals and/or abstract objects, theories of mathematical knowledge and mathematical objects, theories of propositions, Russell’s theory of definite descriptions, three-dimensionalism and endurantism, and deflationism. Instructors may include some of those topics in their teaching of the book if they want.

This book stemmed from my teaching of a course titled “Philosophical Concepts” at Virginia Commonwealth University in Fall 2017, 2018, and 2019. The then-chair of our philosophy department Tony Ellis and I discussed the need for a course like this, and Tony encouraged me to try out the idea. I would like to thank Tony for his encouragement, as well as the students who participated in the course for their input. I also would like to thank Rod Bertolet, Chad Carmichael, Michael Nelson, Katharine Schweitzer, Jim Scow, Tomoji Shogenji, and Cathy Sutton for reading earlier manuscripts

of the book and giving me valuable advice. Special thanks go to my student assistant Sarah Colquhoun, who carefully read through the penultimate manuscript and gave me numerous suggestions for improvement, and Elizabeth Bolstad, who kindly provided me with nice illustrations. Finally, I dedicate this book to my wife Chong, who made my life full of joy and happiness when I was writing this book.

# Particulars and Universals; Logic and Language

---

### 1.1 Tokens and Types; Particulars and Universals

To begin, let's take a look at the whiteboard (Figure 1.1).

How many words are there on the board? Two or three? Some of you may say that there are two words on the board, 'dog' and 'cat'. Others may say that there are three words, 'dog', 'cat', and another 'dog' at the bottom. But whether you say 'two' or 'three', you'd agree that there is a sense in which the other answer is also correct. So how do you precisely describe the number of words on the board?

This is how. We introduce two terms, a *word token* and a *word type*. A word token is a particular inscription or utterance of a word, whereas a word type is a type of word which can be expressed possibly by multiple word tokens. So, in the present example, there are three word tokens 'dog', 'cat', and another 'dog', written on the board, expressing two word types, 'dog' and 'cat'.

This type/token distinction can be used not only for words but more generally. For instance, philosophers talk about *event tokens* and *event types*. If I kiss my wife twice, each kissing can count as an event token; so there are two event tokens here, two kissings. But there is one event type, kissing.

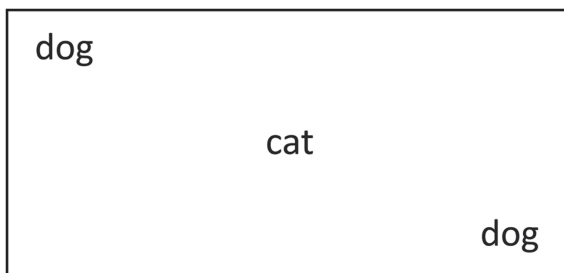


Figure 1.1 Words on the whiteboard.

(Actually, those two event tokens can be tokens of a different event type, including *kissing*, *kissing a woman*, *kissing somebody's wife*, and *kissing Ken's wife*.)

The term 'token' should be understood as a technical term. You may have heard the word in a totally different context before. For instance, a special kind of coin you purchase and insert into the slot of a turnstile at a subway station may be called 'token' (though nowadays they have mostly been replaced with Metrocards). If someone says 'as a token of my appreciation', that 'token' basically means 'symbol'. Anyway, take the word 'token' introduced here as a special technical term.

Then, how would you describe the situation in my backyard (Figure 1.2)? How many animals are there?

If you say, in analogy with the last example, that there are three animal tokens, two dogs and one cat, expressing two animal types, dog and cat, that's absolutely correct. There is nothing wrong with that. Philosophers, however, also traditionally use another pair of technical terms: *particulars* instead of 'tokens' and *universals* instead of 'types'. Then there are three particulars, two dogs and one cat, Max, Fido, and Tibbles, in the backyard; they *instantiate* (or *exemplify*) two universals, dogness and catness. They also instantiate the universal animalness. I define a universal as a thing (in the broadest sense, or an *entity*, as many philosophers put it) that can be instantiated by particulars; more specifically, it can be instantiated by multiple particulars at different locations at once. Particulars are also called *individuals* in this book.

Thus, the particulars Max and Fido instantiate the universal dogness, the particular Tibbles instantiates the universal catness, and they all instantiate animalness. Considered the other way around, we *abstract* dogness, catness, and animalness from those particulars. So universals are a species of *abstract objects*.<sup>1</sup> Generally, an abstract object is an object that has no spatiotemporal location; it exists outside of spacetime. Objects that are not abstract are *concrete objects*; so a concrete object exists somewhere in spacetime.

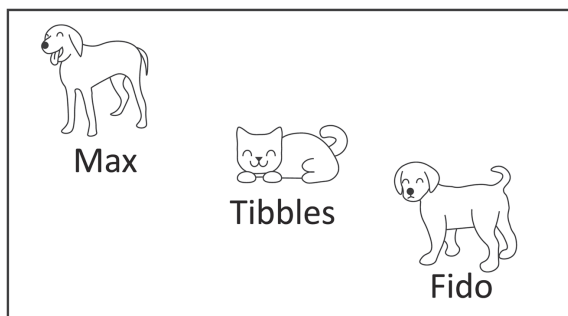


Figure 1.2 Animals in my backyard.

Can you give examples of particulars and universals? Particulars are easy to find: all individual people, such as Barack Obama, Donald Trump, Queen Elizabeth II, each of you, me, etc.; individual physical objects such as this table, that chair, that building, etc.; tiny objects such as cells, molecules, atoms, and fundamental particles and large objects such as New York City, the USA, Earth, and the Milky Way galaxy. Note that each individual may have *parts*, which themselves are individuals. So, for instance, the Milky Way galaxy contains 200 billion solar systems including ours; our solar system contains the Sun and eight planets including Earth; and Earth contains a few hundred countries, each of which contains many cities and towns, etc. We will talk more about parts later, but all those parts can be considered particulars of their own right.

Universals are, or at least include, properties and relations shared by particulars. They include *properties* such as beauty, honesty, courage, tallness, shortness, roundness, squareness, blackness, whiteness, redness, greenness, humanness (or humanity), dogness, catness, etc. (As you can see, if you have an adjective or a common noun such as ‘tall’ or ‘dog’ and add ‘-ness’ at the end, most likely the result will be a name of a property.) There are also such properties as *being an old professor*, *being a bright student*, *being a black chair*, and *being a round table*. (Then tallness = the property *being tall*; dogness = the property *being a dog*, etc.) Some properties are instantiated by particulars for a long period of time; others are instantiated only briefly. So two young people jogging instantiate the property *jogging* (or joggingness?) only so long as they are jogging, but they instantiate the property *youth* for several years and *personhood* so long as they live as persons.

While properties are instantiated by single individuals, *relations* are instantiated by groups of individuals. Examples are love, hate, kissing, kicking, speaking to, the relations *being taller than*, *being heavier than*, *sitting next to*, etc. These are *2-place relations*; that is, each of them is instantiated by *pairs* of particulars. There are also *3-place*, *4-place*, etc., *relations*. For instance, *giving* ( $x$  gives  $y$  to  $z$ ) and *being between* ( $x$  is between  $y$  and  $z$ ) are 3-place relations. Just like properties, relations can last for a long time or only briefly. Two students can be sitting next to each other only during the class meeting; two people can be in love with each other for a very long time or a relatively short time. When the term ‘property’ is used broadly, it includes relations as well as properties in the narrow sense as above.

I said that universals are species of abstract objects, i.e., objects that exist outside of spacetime. One crucial difference between particulars and universals is that particulars have spatiotemporal locations: they are located somewhere in spacetime. In the above example, Max and Fido are standing at the two corners of the backyard this afternoon, and Tibbles is sitting at the center. In contrast, universals do not have specific spatiotemporal locations; they do not exist in spacetime. Dogness is not located either at the top left corner of the yard or at the bottom right corner. Analogously, catness is not

located at the center of the yard even though there is only one particular cat in the yard that instantiates catness and she is located at the center. If there is another cat outside of the yard, then she instantiates catness, too. One important common feature of universals is that they all exist outside of spacetime.<sup>2</sup>

Do all particulars exist in spacetime? That's a good question we won't answer here. All physical objects seem to exist in spacetime. But how about numbers, 1, 2, 3 ..., for instance? They seem to be abstract objects and do not seem to exist in specific locations in spacetime. Number 2, the number of dogs in the backyard, does not seem to exist in the backyard. Are the numbers particulars or universals? That's a good but difficult question we won't try to answer here. But there is no question that typical particulars like material objects do have spacetime locations.

Those who studied ancient philosophy probably know Plato's Theory of Forms. According to that theory, Forms are essences of individuals which exist on their own. For instance, separate from this beautiful woman and that beautiful man, there exists a Form of *beauty*. Each beautiful person is only an imperfect instantiation of the Form, which is perfect. Two beautiful people are beautiful by virtue of both (imperfectly) instantiating the same Form. The world of individuals is an imperfect, constantly changing world whereas the world of Forms is a perfect, never-changing world, according to Plato.

Plato's Forms are the earliest examples of universals. Thus the view that there are universals existing independently of particulars is called *platonism* (often with the lowercase 'p').<sup>3</sup> The opposite of platonism is *nominalism*. 'Nominal' here means 'name only' (e.g., 'a *nominal* gratuity'). So nominalism denies the existence of universals and holds that words such as 'beauty' and 'honesty' are names only and do not stand for any actual objects. For the sake of discussion, in what follows we will assume that platonism is correct, and that there are universals (and abstract objects) in reality.

The study or theory of what exists and what not is called *ontology*. Ontology is a part of *metaphysics*, which is the study or theory of the existence and general nature of things in reality. So platonism and nominalism are ontological theories about universals.

## 1.2 Realism and Anti-realism

To put the distinction between platonism and nominalism into perspective, let's introduce the general distinction between *realism* and *anti-realism*. For any object *x* or group *x* of objects, *realism* about *x* is the view that *x* exists. Its denial, the view that *x* does not exist, is *anti-realism* about *x*. So, for instance, realism about the external world holds that the world exists outside of our minds, whereas anti-realism about the external world, like George Berkeley's view, claims that there is no external world. Realism about morality, or *moral*



*realism*, is the view that moral properties such as *being morally right* and *being morally wrong* exist objectively, independently of what we think about them, instantiated by some actions but not others; so there are objective answers to moral questions such as what is morally right and what is morally wrong. Moral anti-realism denies the existence of moral properties and moral facts. (We will discuss moral anti-realism later in Sections 4.7 and 4.9.) Realism about mathematical objects, or *mathematical realism*, holds that mathematical objects such as numbers and sets exist (presumably outside of spacetime), while mathematical anti-realism denies it. Generally, realism and anti-realism are ontological theories.

Anti-realism about  $x$  is sometimes called *nihilism* about  $x$ , where ‘nihil’ means ‘non-existence’. If you look up ‘nihilism’ online or in a dictionary, you will find a definition like ‘the view that life is without meaning or intrinsic value’ and find the names of some nineteenth-century philosophers such as Friedrich Nietzsche and Søren Kierkegaard associated with it (though my trusted colleagues assure me that they are not nihilists even in that sense). Nihilism in our sense is much more general than nihilism in this sense. A version of anti-realism about  $x$  that maintains that  $x$  is a useful fiction is called *fictionalism* about  $x$ . So, for instance, mathematical fictionalism holds that mathematical objects such as numbers and sets do not really exist, but that it is useful (e.g., for scientific purposes) to talk about them as if they existed.

In this terminology, we can say that platonism and nominalism are realism and anti-realism about universals (and abstract objects), respectively.

### 1.3 Propositional Logic

Language is a major topic in contemporary philosophy. We will discuss various issues pertaining to language throughout this book. Our analysis of language is based on that given in *propositional logic* and *predicate logic*. I assume in this book that you have already taken at least one course in logic which includes propositional logic but not necessarily predicate logic. I will sketch the relevant part of propositional and predicate logic in this section and the next. You should come to understand the distinction between particulars and universals more clearly afterwards.

Generally speaking, logic is the study and theory of *valid arguments*. An argument consists of two parts, premises and the conclusion, and, informally, an argument is valid if and only if its conclusion logically follows from its premises (or the premises logically imply the conclusion). Slightly more formally:

- Validity  
An argument is *valid* if and only if the following is impossible: all its premises are true and yet its conclusion is false.

However, whether an argument in natural language (such as English) is valid or not depends on how it is formalized; that's where the differences between various logics with different formalizations come into the picture.

Predicate logic, also called *quantificational logic* or *first-order logic*, was invented (or discovered) by Gottlob Frege and Bertrand Russell a little more than a century ago. It is an extension of a simpler logic called *propositional (or sentential) logic*: every legitimate formula or inference rule in propositional logic is also a legitimate formula or inference rule in predicate logic, but there are formulas and inference rules that are legitimate only in predicate logic.

In propositional logic, the simplest meaningful expressions are *atomic sentences*,  $P, Q, R, \dots$  (Or atomic *propositions*. However, we will avoid using the word 'propositions' in this context because we will use the word differently in the subsequent. Simply put, on our usage propositions are not themselves sentences but the *meanings* of declarative sentences.) Atomic sentences can be connected by *logical connectives* to form *complex (or compound) sentences*. There are five connectives: negation  $\neg$ , conjunction  $\wedge$ , disjunction  $\vee$ , conditional  $\rightarrow$ , and biconditional  $\leftrightarrow$ . For any sentences  $P$  and  $Q$ , *not P*, *P and Q*, *P or Q (possibly both)*,<sup>4</sup> *if P then Q*, and *P iff (= if and only if) Q*, can be symbolized as  $\neg P$ ,  $P \wedge Q$ ,  $P \vee Q$ ,  $P \rightarrow Q$ , and  $P \leftrightarrow Q$ , respectively.<sup>5</sup> So suppose:

- $P$ : Adam is a professor.
- $Q$ : This is a university building.

Then:

- $\neg P$ : Adam is not a professor.
- $P \wedge Q$ : Adam is a professor and this is a university building.
- $P \vee Q$ : Adam is a professor or this is a university building.
- $P \rightarrow Q$ : If Adam is a professor, then this is a university building.
- $P \leftrightarrow Q$ : Adam is a professor if and only if this is a university building.

$P$  and  $Q$  of conjunction  $P \wedge Q$  are called *conjuncts*, and  $P$  and  $Q$  of disjunction  $P \vee Q$  are called *disjuncts*; the if-part  $P$  and the then-part  $Q$  of the conditional  $P \rightarrow Q$  are called *the antecedent* and *the consequent*, respectively. Some sentences are logically related to others. For instance,  $P \wedge Q$  logically implies  $P$  (and  $Q$ );  $P$  (or  $Q$ ) implies  $P \vee Q$ ; and  $P \rightarrow Q$  and  $P$  together imply  $Q$ .

The simplest way to express the meanings of the logical connectives is by means of *truth tables*. The following five tables tell us, for any  $P$  and  $Q$ , when  $\neg P$ ,  $P \wedge Q$ ,  $P \vee Q$ ,  $P \rightarrow Q$ , and  $P \leftrightarrow Q$  are true (T) or false (F) depending on the truth values of  $P$  and  $Q$ :

$P$	$\neg P$
T	F
F	T

$P$	$Q$	$P \wedge Q$	$P \vee Q$	$P \rightarrow Q$	$P \leftrightarrow Q$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	F	T	T	F
F	F	F	F	T	T

Put in words,  $\neg P$  is true iff  $P$  is false;  $P \wedge Q$  is true iff  $P$  and  $Q$  are both true;  $P \vee Q$  is false iff  $P$  and  $Q$  are both false;  $P \rightarrow Q$  is false iff  $P$  is true and  $Q$  is false; and  $P \leftrightarrow Q$  is true iff  $P$  and  $Q$  have the same truth value.

The above truth tables determine the truth tables of more complex sentences. I will show below that  $\neg(P \wedge Q)$  (i.e., not both  $P$  and  $Q$ ) and  $\neg P \vee \neg Q$  (not- $P$  or not- $Q$ ) are *logically equivalent* to each other, and that  $\neg(P \vee Q)$  (neither  $P$  nor  $Q$ ) and  $\neg P \wedge \neg Q$  (not- $P$  and not- $Q$ ) are logically equivalent to each other, by showing that they have the same truth table:

$P$	$Q$	$\neg$	$(P$	$\wedge$	$Q)$	$\neg$	$P$	$\vee$	$\neg$	$Q$
T	T	<b>F</b>	T	T	T	F	T	<b>F</b>	F	T
T	F	<b>T</b>	T	F	F	F	T	<b>T</b>	T	F
F	T	<b>T</b>	F	F	T	T	F	<b>T</b>	F	T
F	F	<b>T</b>	F	F	F	T	F	<b>T</b>	T	F

$P$	$Q$	$\neg$	$(P$	$\vee$	$Q)$	$\neg$	$P$	$\wedge$	$\neg$	$Q$
T	T	<b>F</b>	T	T	T	F	T	<b>F</b>	F	T
T	F	<b>F</b>	T	T	F	F	T	<b>F</b>	T	F
F	T	<b>F</b>	F	T	T	T	F	<b>F</b>	F	T
F	F	<b>T</b>	F	F	F	T	F	<b>T</b>	T	F

The boldfaced columns are the truth tables for the whole sentences. The following is the way to read the first row of the first table: Suppose  $P$  is T and  $Q$  is T; then, on the left,  $P \wedge Q$  is T, so  $\neg(P \wedge Q)$  is F; on the right,  $\neg P$  is F and  $\neg Q$  is F; so  $\neg P \vee \neg Q$  is F. Similarly for the other rows. The equivalence between  $\neg(P \wedge Q)$  and  $\neg P \vee \neg Q$  and that between  $\neg(P \vee Q)$  and  $\neg P \wedge \neg Q$  are called *De Morgan's Laws*.

The conditional  $\rightarrow$  we use in logic is called *the material conditional*. As I will show later in Section 3.13, the material conditional is different from two kinds of English conditional, *the indicative conditional* and *the subjunctive conditional*. The material conditional  $P \rightarrow Q$  is logically equivalent to  $\neg P \vee Q$ , as the following truth table shows:

$P$	$Q$	$P$	$\rightarrow$	$Q$	$\neg$	$P$	$\vee$	$Q$
T	T	T	<b>T</b>	T	F	T	<b>T</b>	T
T	F	T	<b>F</b>	F	F	T	<b>F</b>	F
F	T	F	<b>T</b>	T	T	F	<b>T</b>	T
F	F	F	<b>T</b>	F	T	F	<b>T</b>	F

## 1.4 Predicate Logic

A major shortcoming of propositional logic is that in it, the (atomic) sentences are the smallest meaningful units and that their internal structures are completely ignored. For instance, ‘Adam is tall’ and ‘Betty is a professor’ have obvious logical connections to ‘Adam is a professor ( $P$ )’ which ‘This is a university building ( $Q$ )’ does not have, but in propositional logic that difference is ignored and ‘Adam is tall’ and ‘Betty is a professor’ are given symbolizations totally unrelated to  $P$  or  $Q$ , such as  $R$  and  $S$ .

This is where predicate logic becomes useful. Each atomic sentence can be separated into two constituents, *singular terms* and a *predicate*. Singular terms refer to single objects (particulars or individuals), and predicates stand for universals. Predicates are divided into 0-place predicates (= declarative sentences), 1-place predicates, 2-place predicates, 3-place predicates ..., which stand for, respectively, propositions, properties, 2-place relations, 3-place relations .... Singular terms are symbolized as lowercase letters,  $a$ ,  $b$ ,  $c$  ..., and predicates, e.g.,  $Px$ ,  $Qxy$ ,  $Rxyz$  ..., are uppercase letters followed by variables,  $x$ ,  $y$ ,  $z$  .... For any  $n$  ( $\geq 0$ ), an  $n$ -place predicate involves  $n$  distinct variables. If  $n$  is 2 or larger, the predicate is called a *many-place* predicate.

Note that understood this way, propositions are universals analogous to properties and relations. Properties and relations are universals instantiated by individuals and groups of individuals, respectively; propositions, then, are extreme cases of universals that need no individuals to instantiate. Recall that, even though predicates may stand for (represent, or express) properties and relations, properties and relations, as universals, exist independently of, and prior to, language. Generally, universals exist independently of us, humans, and the language we speak. Consequently, propositions also exist independently of us or the declarative sentences we assert that may express them.

Let me give you a few examples of how to combine singular terms and predicates to produce sentences. Suppose  $a$  is the abbreviation of ‘Adam’ and refers to Adam;  $b$  is the abbreviation of ‘Betty’ and refers to Betty;  $Px$  is the abbreviation of ‘ $x$  is a professor’ and stands for the property *being a professor* (or professorship);  $Tx$  is the abbreviation of ‘ $x$  is tall’ and stands for the property *being tall* (or tallness); and  $Lxy$  is the abbreviation of ‘ $x$  loves  $y$ ’ and stands for the 2-place relation  $x$ ’s *loving*  $y$ . Then, for instance:

- $Pa$ : Adam is a professor.
- $Pb$ : Betty is a professor.
- $Ta$ : Adam is tall.
- $\neg Tb$ : Betty is not tall.
- $Lab$ : Adam loves Betty.
- $\neg Lba$ : Betty does not love Adam.
- $Lab \wedge \neg Lba$ : Adam loves Betty, but Betty does not love Adam.

We can also turn an  $n$ -place predicate into an  $(n - 1)$ -place predicate (i.e., a 3-place predicate into a 2-place predicate, a 2-place predicate into a 1-place predicate, a 1-place predicate into a 0-place predicate, i.e., a declarative sentence, etc.) by filling in one of the variables ( $x, y, z$ , etc.) with a singular term. For instance, we can turn the 3-place predicate  $Gxyz$  (' $x$  gives  $y$  to  $z$ ') into the 2-place predicate  $Gxyb$  (' $x$  gives  $y$  to Betty') by filling in the variable  $z$  with the singular term  $b$ , or 'Betty'; we can turn the 2-place predicates  $Lxy$  (' $x$  loves  $y$ ') and  $Gxyb$  (' $x$  gives  $y$  to Betty') into the 1-place predicates  $Lxb$  (' $x$  loves Betty') and  $Gxtb$  (' $x$  gives Tibbles to Betty') by filling in the variable  $y$  with 'Betty' and 'Tibbles', respectively; and we can turn the 1-place predicates  $Tx$  (' $x$  is tall') and  $Lxb$  (' $x$  loves Betty') into the declarative sentences  $Ta$  ('Adam is tall') and  $Lab$  ('Adam loves Betty') by filling in the variable  $x$  with 'Adam'.

There is another group of important devices in predicate (or 'quantificational') logic: the *universal quantifier*,  $\forall x$  ('for any  $x$ '), and the *existential quantifier*,  $\exists x$  ('for some  $x$ ').<sup>6</sup> However, we will not discuss them in detail in this book because, by and large, we don't need them (with a few exceptions). Let me just give a few symbolizations and their interpretations to indicate how the quantifiers work:

- $\forall xTx$  (i.e., for any  $x$ ,  $Tx$ ): Everything is tall.
- $\exists xPx$  (for some  $x$ ,  $Px$ ): Something is a professor.
- $\forall x(Px \rightarrow Tx)$  (for any  $x$ , if  $Px$  then  $Tx$ ): Every professor is tall.
- $\exists x(Px \wedge Tx)$  (for some  $x$ ,  $Px$  and  $Tx$ ): Some professor is tall.

A set of important logical laws involving the quantifiers is:

- Generalized De Morgan's Laws:  
For any 1-place predicate  $Px$ ,
  - $\neg \forall xPx$  (not every  $x$  is  $P$ ) iff  $\exists x\neg Px$  (some  $x$  is not- $P$ );
  - $\neg \exists xPx$  (no  $x$  is  $P$ ) iff  $\forall x\neg Px$  (every  $x$  is not- $P$ ).<sup>7</sup>

Consequently, the universal quantifier and the existential quantifier are mutually definable:  $\forall xPx$  (every  $x$  is  $P$ ) =  $\neg \exists x\neg Px$  (no  $x$  is not- $P$ ) and  $\exists xPx$  (some  $x$  is  $P$ ) =  $\neg \forall x\neg Px$  (not every  $x$  is not- $P$ ). So we don't need both quantifiers as primitives; if we have one, we can introduce the other by definition. But that's all for quantifiers.

Below I summarize the types of sentences, singular terms, and predicates. You should read the following summary very carefully; there is much content in it.

- Sentences
  - Declarative sentences: You came here to study. She runs every day. ...
  - Interrogative sentences: Did you come here to study? Does she run every day? ...
  - Imperative sentences: Come here. Run every day. Open the door. ...
  - Exclamatory sentences: Oh! Darn it! Hooray! Boo! Gee! ...
- Singular terms
  - Proper names: Adam, Betty, Charlie, Max, Fido, Tibbles, Donald Trump, Elizabeth II, Pegasus, Santa Claus, Mt Everest, New York City, Virginia, the United States of America, the Fountain of Youth, Atlantis, Earth, the Sun, the Milky Way galaxy ....
  - Definite descriptions (often ‘the + such and such (singular)’): the current President of the United States, the present King of France, the largest prime number, the son of Joshua, Joshua’s son, mean molecular kinetic energy, the highest mountain in the world, the best student in class, the student sitting in front of me, the tall girl, the round table, the chair ....
  - Demonstratives: this, that, these, those, I, you ....
- Predicates
  - 0-place predicates (= declarative sentences): It is raining. It is cold. ...
  - 1-place predicates:
    - Adjectival predicates (predicates whose cores are adjectives):  $x$  is beautiful,  $x$  is honest,  $x$  is courageous,  $x$  is tall,  $x$  is short,  $x$  is round,  $x$  is square,  $x$  is black,  $x$  is white,  $x$  is young,  $x$  is old,  $x$  is male,  $x$  is female,  $x$  is human ....
    - Nominal predicates (predicates whose cores are common nouns):  $x$  is a male,  $x$  is a female,  $x$  is a human,  $x$  is a teacher,  $x$  is a professor,  $x$  is a dog,  $x$  is a cat,  $x$  is an animal,  $x$  is a table,  $x$  is a chair,  $x$  is a round table,  $x$  is a better student in class,  $x$  is a student who is taking a logic course ....
    - Verbal predicates (predicates whose cores are intransitive verbs):  $x$  walks,  $x$  runs,  $x$  talks,  $x$  sleeps ....
    - Adverbial predicates (predicates whose cores are adverbs):  $x$  is abroad,  $x$  is inside ....
  - 2-place predicates:  $x$  loves  $y$ ,  $x$  hates  $y$ ,  $x$  kisses  $y$ ,  $x$  kicks  $y$  (grammatically, these are transitive verbs),  $x$  is identical with  $y$  ( $x = y$ ),  $x$  is taller than  $y$ ,  $x$  is heavier than  $y$ ,  $x$  is next to  $y$ ,  $x$  is in front of  $y$  ....

- 3-place predicates:  $x$  gives  $y$  to  $z$ ,  $x$  hands  $y$  to  $z$ ,  $x$  reminds  $y$  of  $z$ ,  $x$  is between  $y$  and  $z$ , ....
- $n$ -place predicates ....

Throughout this book, when we talk about linguistic expressions, we will come back again and again to these three kinds of expressions: singular terms, predicates, and sentences. More specifically, we will focus on singular terms and 1-place predicates, as well as *declarative* sentences as combinations of singular terms and 1-place predicates. We will often omit the words '1-place' and 'declarative' and simply say 'predicates' and 'sentences' when we are talking about 1-place predicates and declarative sentences. As for singular terms, we will focus on the first two types of singular terms in the list, proper names and definite descriptions,<sup>8</sup> setting aside demonstratives. We will use the term '(linguistic) expression' as a more general term that includes singular terms, predicates, and sentences.

The above list is mostly self-explanatory, so I have little to add, but I would like to bring two important related facts to your attention. First, common nouns, such as 'dog', 'cat', and 'professor', are very different from proper names (or 'proper nouns'), such as 'Max', 'Tibbles', and 'Adam'; the latter are singular terms, while the former are part of 1-place predicates ' $x$  is a dog', ' $x$  is a cat', ' $x$  is a professor', etc.

Second, indefinite descriptions, 'a(n) (i.e., indefinite article) + such and such' (where 'such and such' is a common noun phrase), such as 'a President of the United States', 'a son of Joshua', and 'a better student in class', are very different from definite descriptions of the form 'the (i.e., definite article) + such and such', such as 'the President of the United States', 'the son of Joshua', and 'the best student in class'; the latter are singular terms, while the former are part of 1-place predicates. For instance, 'the best student in class' is a singular term (in particular, a definite description), referring to the best individual in class, whoever it is. In contrast, 'a better student in class' is part of the 1-place predicate ' $x$  is a better student in class'; similarly, 'the teacher of Alexander the Great' is a singular term (a definite description) while 'a teacher' (or even 'a teacher of Alexander the Great') is part of the 1-place predicate ' $x$  is a teacher (of Alexander the Great)'.<sup>9</sup> It is interesting to realize that what at first glance seems like a small difference (proper vs common noun; definite vs indefinite article) turns out to make a huge difference after analysis.

We've gone a great length to spell out the distinctions in sentences, singular terms, and predicates. That's partly because we will come back to those distinctions many times in this book. The point I would like to make at this point, however, is a simple one: the crucial distinction in predicate logic between singular terms and predicates closely corresponds to the traditional distinction between particulars and universals; more specifically, singular terms and predicates should be considered to be expressions for particulars

and universals, respectively, and the sentence ‘ $Pa$ ’ (e.g., ‘Adam is a professor’ or ‘Fido is a dog’) is true if and only if the particular denoted<sup>10</sup> by ‘ $a$ ’ (Adam, Fido, etc.) instantiates the property denoted by ‘ $Px$ ’ (*being a professor, being a dog*, etc.).<sup>11</sup> The sentence is false if and only if the particular denoted by ‘ $a$ ’ does not instantiate the property denoted by ‘ $Px$ ’. Note that there are singular terms, such as ‘Pegasus’, ‘the Fountain of Youth’, ‘Atlantis’ (proper names), ‘the present King of France’, and ‘the largest prime number’ (definite descriptions), that do not denote any actually existing thing. We will discuss how to deal with those ‘empty’ singular terms later.

To summarize:

Singular terms <sup>12</sup>	stand for (represent, or express) $\Rightarrow$	Particulars (individuals)
Predicates	$\Rightarrow$	Universals
– 1-place predicates	$\Rightarrow$	– Properties
– Many-place predicates	$\Rightarrow$	– Relations
(Declarative) sentences (= 0-place predicates)	$\Rightarrow$	– Propositions

Note that this is still an incomplete list – very incomplete, indeed. We will see in the next chapter that the linguistic expressions on the left of the list stand for not just one kind of thing, as the list shows, but two kinds (which we call ‘extension’ and ‘intension’). For instance, we already saw that declarative sentences also have truth values, Truth and Falsity. What’s the relation between propositions and truth values? The answer will be: declarative sentences denote truth values as their extensions but connote propositions as their intensions. But that’s for later.

Predicate logic was invented (or discovered) by Gottlob Frege and Bertrand Russell only a little more than a century ago, while platonism has been around for more than twenty-five centuries. It is always nice to see a connection between something very old and something rather new (relatively speaking, of course).

1.5 Identity

In the last two sections we have had a general discussion about propositional and predicate logic. In the rest of this chapter, I would like to give a little more detail to some of the relevant topics: identity, necessary and sufficient conditions, and quotation. This section deals with the first topic: identity.

I’ve said above that we will focus on 1-place predicates, but one exception to this policy is the 2-place predicate of identity (or equality):  $x = y$  ( $x$  is identical with  $y$ , or  $x$  equals  $y$ ). The identity predicate is very important both philosophically and mathematically, and it has its own distinctive logic.



Because of that, it is often considered a logical operator along with the connectives and quantifiers. The version of predicate logic which involves the identity predicate as a logical operator is called *predicate logic with identity*. Our later discussions will sometimes involve the identity predicate. So I would like to spell out the logic of the identity predicate in this section. What the identity predicate stands for is so-called *numerical (or strict) identity*. It is also very important to distinguish numerical identity from other, looser kinds of identity we often talk about in both philosophical and non-philosophical contexts; so I will also present a few examples of identity which are not numerical identity.

But, first, numerical identity. It is often said that everything is identical with itself and nothing else. Identity in this sense is numerical identity. Two things can never be numerically identical because two things cannot be one. The teacher of Alexander the Great and the most famous disciple of Plato are numerically identical because they are Aristotle, a single person. The teacher of Alexander the Great and the most famous disciple of Socrates are numerically non-identical (or ‘distinct’) because the former is Aristotle while the latter is Plato, two distinct persons. ‘Identical’ generally means ‘same’; ‘identical’ and ‘same’ mean the same (identical) thing. But for a numerically identical thing, we tend to say ‘one and the same<sup>13</sup> thing’, emphasizing the fact that we are talking about one thing, not two or more things. Some people tend to use ‘same’ instead of ‘identical’ when they are talking about identity of looser kinds (which I will introduce shortly). But that’s not a strictly enforced usage; ‘identical’ can still be used for identity of looser kinds, too. It is customary to drop the word ‘numerical’ or ‘strict’ and simply say ‘identity’ when we are talking about numerical (or strict) identity. We employ this convention throughout this book.

In mathematics, identity is called ‘equality’, but it is the same (identical) thing. Mathematics books and papers are full of equations, which are really statements of identity. For example,  $2+3 = 10/2$ ; both sides of this equation stand for one and the same thing, number 5.

Let me give you another example of numerical identity. We will come back to this example many times in this book. It is a very famous example among philosophers. The two philosophers whose names will be most often mentioned in the rest of this book, the aforementioned German logician-philosopher Gottlob Frege and the contemporary American logician-philosopher Saul Kripke, both used this example and made it famous. ‘Hesperus’ (or ‘the Evening Star’ in English) is the name of the bright heavenly body you see above the west horizon just after the sunset. ‘Phosphorus’ (or ‘the Morning Star’) is the name of the bright heavenly body you see above the east horizon just before the sunrise. We lose track of these heavenly bodies during the daytime because the sunlight is too strong. We discovered at one point of history, however, that Hesperus is identical with Phosphorus, which is really the planet Venus (not a star). Here Hesperus

and Phosphorus are one and the same thing we see at different times of the day. This is another example of numerical identity.

Logically speaking, the identity relation is an equivalence relation. What is an equivalence relation? An *equivalence relation* is any 2-place relation  $R$  which has the following three properties:

- The equivalence relation  
For any objects  $x$ ,  $y$ , and  $z$ ,
  - Reflexivity:  $Rxx$ .
  - Symmetry: If  $Rxy$ , then  $Ryx$ .
  - Transitivity: If  $Rxy$  and  $Ryz$ , then  $Rxz$ .

The group of things that are in the equivalence relation  $R$  is called *the equivalence class* (with respect to  $R$ ). Since the identity relation is an equivalence relation, the following holds for (numerical) identity =:

- (Numerical) identity  
For any objects  $x$ ,  $y$ , and  $z$ ,
  - Reflexivity:  $x = x$ .
  - Symmetry: If  $x = y$ , then  $y = x$ .
  - Transitivity: If  $x = y$  and  $y = z$ , then  $x = z$ .

However, there are many equivalence relations that are not the identity relation; an equivalence class may have more than one member. For instance,  $x$  *has the same biological mother as*  $y$  and  $x$  *has the same height as*  $y$  are equivalence relations but obviously not the identity relation between  $x$  and  $y$ . I've touched on *logical equivalence* in Section 1.3. Logical equivalence is a species of equivalence relations, but two sentences' being logically equivalent does not mean that they are not two sentences but one. What distinguishes the identity relation from the other equivalence relations is the following law attributed to the great philosopher-mathematician Gottfried Leibniz:

- Leibniz's Law (= the Indiscernibility of Identicals + the Identity of Indiscernibles)  
For any objects  $x$  and  $y$ ,  
 $x = y$  iff  $x$  and  $y$  have exactly the same properties.<sup>14</sup>

The 'only if' direction ( $\rightarrow$ ) of the Law is called *the Indiscernibility of Identicals*, and the 'if' direction ( $\leftarrow$ ) of the Law is called *the Identity of Indiscernibles*.<sup>15</sup> The former, which states that  $x$  and  $y$  are distinct if they have at least one different property, comes from the concept of identity. The latter is derivable from the concept of identity that includes the reflexivity and symmetry of the identity relation. To derive it, assume the right-hand side of the biconditional above, i.e., that  $x$  and  $y$  have exactly the same properties. Since  $x$  is identical with  $x$ ,

$x$  has the property *being identical with  $x$* . Thus,  $y$  must have that property. Thus,  $y$  is identical with  $x$ . By symmetry,  $x$  is identical with  $y$ . QED.<sup>16</sup>

Numerical identity is really a simple thing, but identity can be a philosophical problem because there are other kinds of identity than numerical identity which are more problematic, and also because they can be confused with numerical identity.

For instance, we may say, 'This soda can is identical with that soda can – the same logos, the same pictures, and the same prints with the same expiration dates'. We are not talking about numerical identity of cans here; we are not saying that this soda can and that soda can are somehow not two distinct cans but a single can. Similarly, if we say about Figure 1.1 before, 'The word token at the top left and the word token at bottom right are identical (i.e., 'dog')', we are not talking about a single word token but two distinct word tokens. In these cases, 'identical with' means 'of the same kind as'. This soda can is of the same kind as that soda can, and this word token is of the same kind as that word token (the kind being a universal). If your dorm room and your friend's dorm room are identical (i.e., the same layout, the same wall color, etc.), that does not mean that you and your friend live in one and the same doom room.

There is a little complication here, however. This soda can and that soda can (this word token and that word token, or this dorm room and that dorm room) instantiate *the one and same type* (i.e., universal). Then we are talking about numerical identity of types (universals). So you have to be careful about what you are talking about. In Section 1.1, we drew the distinction between tokens and types. If potentially two tokens turn out to be one and the same token (individual, or particular), we are dealing with *token identity*. On the other hand, if potentially two types turn out to be one and the same type (universal), then we are talking about *type identity*. Aristotle and the most famous disciple of Plato are token identical; so are Mark Twain and Samuel Clemens. Lightning and atmospheric electrical discharge, or, more precisely, the property of *being a lightning* and that of *being an atmospheric electrical discharge*, are type identical; so every lightning is an atmospheric electrical discharge and vice versa. Similarly, pain and c-fiber firing in the brain, or the property of *being a pain* and that of *being a c-fiber firing*, are type identical; so everything that is a pain is a c-fiber firing and vice versa.

But, again, we do talk about identity that is not numerical identity, such as the identity between this soda can and that soda can. I have already given you in this section the famous Hesperus/Phosphorus example for numerical identity, but let me give you here another famous – probably even more famous – example involving identity – but, this time, not numerical identity but identity of a different kind. It is called 'the Ship of Theseus'. This puzzle has been discussed by many philosophers since the ancient times. The Greek hero Theseus's Ship is kept in a harbor as a commemoration. As it loses its planks one by one due to loose nailing, new planks – exact replicas of the

original planks – are placed to the original places. At the end, all the original planks ended up being replaced with new planks. Let's call the resulting ship 'the Repaired Ship'. At the same time, each of the lost original planks is kept and placed in exactly the same place as the original on land at a museum, and at the end, we have a ship whose planks match exactly the original. Call this ship 'the Reassembled Ship'. The puzzle is: Which of the two ships is the same ship as the Original Ship of Theseus, the Repaired Ship or the Reassembled Ship? Perhaps both? Perhaps neither? Note that, were it not for the Reassembled Ship, we would have said that no doubt, the Repaired Ship is the same ship as the Original Ship of Theseus; it's just a matter of replacing parts. And were it not for the Repaired Ship, we would have said that no doubt, the Reassembled Ship is the same ship as the Original Ship; it's just a matter of disassembling, moving, and reassembling a ship. But what will happen if those two operations are performed simultaneously?

This is a fun puzzle to think about. I do not intend to give you any answer here. (Sorry!) Instead, I would like to say the following in connection with our current topic, identity: assuming the theory called *four-dimensionalism*, the 2-place relation *x is the same (or identical) ship as y* in question here is *not* that of numerical identity.

There are two competing major theories of spatiotemporal objects: one is *three-dimensionalism* and the other is *four-dimensionalism*. Put simply, three-dimensionalism holds that material objects such as chairs, tables, and humans are three-dimensional spatial objects existing through time, whereas four-dimensionalism holds that material objects are four-dimensional objects existing in the four-dimensional manifold, spacetime. In this book I will embrace four-dimensionalism and set aside three-dimensionalism because four-dimensionalism is much easier for the reader to understand. According to four-dimensionalism, (at least many) material objects like us stretch out in the temporal dimension just as a sausage stretches out in one spatial dimension (Figure 1.3). For instance, if I (Ken Akiba) was born in 1960 and

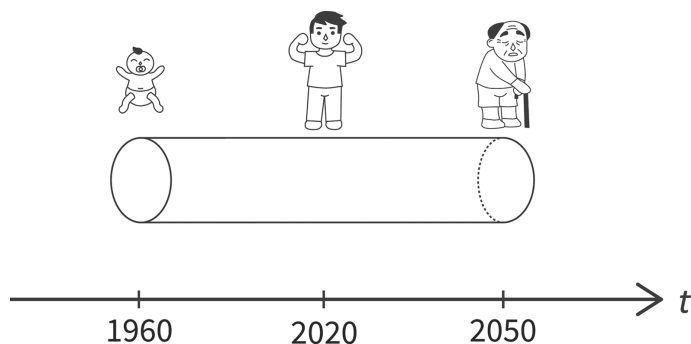


Figure 1.3 A four-dimensional object.

will die in 2050, then Ken Akiba has a fourth, temporal, dimension which stretches out for 90 years.

Now, let's reconsider the Ship of Theseus puzzle on the basis of four-dimensionalism (Figure 1.4). Let's define the Original Ship of Theseus as the ship existing until time  $T_1$ , when it loses its first plank, and the Repaired Ship as the ship beginning to exist at the harbor at  $T_2$  ( $>T_1$ ), when the last of the original planks is replaced with a new plank, and the Reassembled Ship as the ship beginning to exist on land at a museum at  $T_2$ , when the last of the original planks is placed to the new creation. (We ignore the possible time gap between the time when the Repaired Ship is created and the time when the Reassembled Ship is created.) Then: Is the Original Ship the same ship as the Repaired Ship? Is the Original Ship the same ship as the Reassembled Ship?

When we rephrase the question about the identity between the ships this way, it should be clear that we are not talking about numerical identity. For if the identity in question were numerical identity, the answer would be obvious: the Original Ship is not identical with the Repaired Ship or the Reassembled Ship. For, by definition, the Original Ship exists only until  $T_1$ , the Repaired Ship and the Reassembled Ship do not exist until  $T_2$ ; so the Original Ship, the Repaired Ship and the Reassembled Ship are three numerically distinct objects.

The identity at issue here, sometimes called *temporal* (or *diachronic*) identity or *identity over time*, is (at least on the four-dimensional construal) not numerical identity but identity of a different kind. (To signal this, many contemporary philosophers use the term (*temporal*) *persistence* instead.) The question 'Is the Original Ship the same ship as the Repaired Ship?' should be understood as something like 'Is the Original Ship a part of at least one four-dimensional ship that the Repaired Ship is a part of?' (or 'Is there at least one four-dimensional ship the Original Ship and the Repaired Ship are

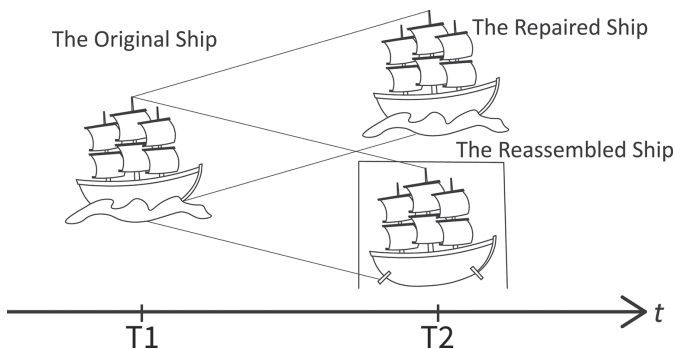


Figure 1.4 The Ship of Theseus.

both parts of?') Whatever the answer to this question may be, what's involved here is not numerical identity.

I've read quite a few reputable philosophers say, against the claim that the Original Ship is the same ship as the Repaired Ship and that the Original Ship is also the same ship as the Reassembled Ship, that this ought to make the Repaired Ship the same ship as the Reassembled Ship by the transitivity of identity (but that that would be untenable). But that does not follow, for we are here not talking about numerical identity, and there is no reason to think that the transitivity must hold also for the kind of identity we are talking about. It is at least logically consistent to hold that the Original Ship is the same ship as the Repaired Ship (i.e., there is a ship both Original Ship and Repaired Ship are parts of), that the Original Ship is the same ship as the Reassembled Ship (i.e., there is a ship both Original Ship and Reassembled Ship are parts of), but that the Repaired Ship is not the same ship as the Reassembled Ship (i.e., there is no single ship both Repaired Ship and Reassembled Ship are parts of). This would make the Original Ship a part of two ships (or two ships overlap in the Original Ship), and some people may not like this consequence; but it is at least logically consistent.

If two ships are in fact overlapping at T1, we cannot tell them apart at T1 even though we can afterwards. Some philosophers introduce the concept of *temporary* (i.e., part-time, not 'temporal') *identity* and say that the two ships are *temporarily identical* at T1 but distinct afterwards. Obviously temporary identity is not numerical identity, as we are talking about two things instead of one. If two things are numerically identical (i.e., not two things but one), then they are *eternally* (and not just temporarily) *identical*; numerical identity implies *eternal identity*.

Generally speaking, *identity over time* is (at least on the four-dimensional construal) not a matter of numerical identity.<sup>17</sup> One popular topic in philosophy is personal identity, and personal identity is a species of identity over time. Consequently, personal identity is not a matter of numerical identity, either. We will discuss personal identity later in Chapter 5, Part C. There (Chapter 5, note 18) we will see some philosophers make the same mistake we discussed above in connection with personal identity.

Identity over time is also a species of non-numerical identity sometimes called *relative identity*. For relative identity, we cannot just ask, 'Is *a* identical with *b*?'; instead we must ask, 'Is *a* the identical *k* with *b*?', where *k* is some *kind* (or *sort*), e.g., ship, person, animal, plant, etc. This is what John Locke (1690, Book II, Section 27) said before he started talking about personal identity. The identity condition differs depending on what kind of things we are talking about. Examples:

- A living creature (animal or plant) growing up may still be the same living creature but is not the same bunch of molecules.

- The Original Ship of Theseus and the Repaired Ship may be the same ship but are not the same mass of lumber.
- On the one hand, if a clay statue of Goliath loses one of its hands, it's still the same statue but a different lump of clay; on the other hand, if the statue is flattened, it's no longer the statue, but is still the same lump of clay. (I will use a modified version of this example later in Section 3.8; so watch out!)
- On the one hand, Adam yesterday and Adam today may be the same human animal but different persons if he totally loses his memory and changes his personality; on the other hand, Adam yesterday and Adam today may be different human animals but the same person if he somehow wakes up in somebody else's body.

Clearly, relative identity is different from numerical identity, for numerical identity is not relative to the kind involved.

To summarize, numerical identity is a simple relation, but there are some other relations we loosely call 'identity', and it is important to distinguish numerical identity and those relations.

Before concluding this section, I should add that the simple 'is' (or, more generally, the 'be' verb, including 'am', 'are', etc.) is often used, instead of the more mouthful 'is identical with (or to)', to signify identity. For instance, instead of saying 'Hesperus is identical with Phosphorus', we can simply say 'Hesperus *is* Phosphorus'. 'Be', used this way as a 2-place predicate of identity =, is called *the 'be' of identity* and is significantly different from 'be' used as part of 1-place adjectival or nominal predicates such as 'is beautiful', 'is tall', 'is (a) human', 'is a professor', and 'is a student' (Section 1.4). 'Be' used in the latter way is called *the 'be' of predication*. So there is a significant difference between the logical structures of 'Adam is the best student in class' and 'Aristotle was the teacher of Alexander the Great', on the one hand, and 'Adam is a better student in class' and 'Aristotle was a teacher (of Alexander the Great)' on the other: the former has the form  $a = b$  while the latter has the form  $Pa$ . This difference is correlated with the difference between definite and indefinite descriptions mentioned in Section 1.4. As you can imagine, there is some resistance to this treatment of 'be'; but this is the most prevalent treatment in linguistics. The third usage of 'be' often mentioned is *the 'be' of existence*. Its examples include: 'There is a cat in my backyard', 'I think, therefore I am', and 'To be or not to be: that is the question'. This 'be' is synonymous to 'exist'; it is an existential quantifier.

## 1.6 Necessary and Sufficient Conditions

Many students in logic are confused about the distinction between necessary conditions and sufficient conditions. They are easily definable in terms of conditionals:

- The condition that  $P$  is a *sufficient condition* for the condition that  $Q$  iff the conditional 'If  $P$  then  $Q$ ' is true.
- The condition that  $Q$  is a *necessary condition* for the condition that  $P$  iff the conditional 'If  $P$  then  $Q$ ' is true.

More simply, suppose 'If  $P$  then  $Q$ ' is true; then (and only then)  $P$  is a sufficient condition for  $Q$ , and  $Q$  is a necessary condition for  $P$ .

This is the way to think about the above definitions: Suppose 'If  $P$  then  $Q$ ' is true. Then if  $P$  is the case, then  $Q$  must be the case. So  $P$  is sufficient for  $Q$  to be the case. Thus,  $P$  is a sufficient condition for  $Q$ . However,  $P$  is not a necessary condition, for even if  $P$  is not the case,  $Q$  may still be the case. On the contrary, if 'If  $P$  then  $Q$ ' is true,  $Q$  must be a necessary concomitant for  $P$ 's being the case:  $P$  cannot be the case without  $Q$ 's being the case, too. So  $Q$  is a necessary condition for  $P$ .

Let's consider a few examples:

- (a) If Fido is a dog, then Fido is an animal. So,
  - The condition that Fido is a dog is a sufficient condition for the condition that Fido is an animal.
  - The condition that Fido is an animal is a necessary condition for the condition that Fido is a dog.
- (b) If today is Christmas Day, then today is in December. So,
  - The condition that today is Christmas Day is a sufficient condition for the condition that today is in December.
  - The condition that today is in December is a necessary condition for the condition that today is Christmas Day.
- (c) If the red (billiard) ball hits the yellow ball, then the yellow ball will move. So,
  - The condition that the red ball hits the yellow ball is a sufficient condition for the condition that the yellow ball will move.
  - The condition that the yellow ball will move is a necessary condition for the condition that the red ball hits the yellow ball.
- (d) If the ground is wet, then it rained last night. So,
  - The condition that the ground is wet is a sufficient condition for the condition that it rained yesterday.
  - The condition that it rained last night is a necessary condition for the condition that the ground is wet.

Some students tend to think that there must be some temporal order between the relevant conditions; for instance, just as a cause always precedes its effects in causation, a sufficient condition for any condition must precede that condition. That's incorrect, however; there are no constraints on the temporal order of necessary and sufficient conditions. As you can see in (a), necessary and sufficient conditions may have nothing



to do with temporal orders; or a sufficient condition for a condition may be simultaneous to (like (b)) or precede (like (c)) or succeed (like (d)) the condition.

Finally, we can define a *necessary and sufficient condition* in the expected way:

- The condition that *P* is a *necessary and sufficient condition* for the condition that *Q* iff the biconditional ‘*P* if and only if *Q*’ is true.

For instance, the condition that Adam is an unmarried adult male is a necessary and sufficient condition for the condition that Adam is a bachelor, and the condition that yesterday was Christmas Eve is a necessary and sufficient condition for the condition that today is Christmas Day.

## 1.7 Quotation

Already in this chapter we have discussed a great deal about the relation between linguistic expressions and the world. The approach to philosophy I and probably 90 percent of philosophers in English-speaking countries, as well as many philosophers in Continental Europe, embrace is called *analytic philosophy*. The two originators of predicate logic, Gottlob Frege and Bertrand Russell, may also be considered the originators of analytic philosophy. The name ‘analytic philosophy’ came from ‘the analysis of language’ and the idea that good philosophy is done by analyzing (i.e., methodically dissecting) language. Russell’s analysis of definite descriptions in “On Denoting” (1905) is a premier exemplification of this idea, and A. J. Ayer’s influential book *Language, Truth, and Logic* (1936) is written from the same viewpoint. Nowadays analytic philosophers are much less stringent in their methodology; but, again, language still plays a large part in contemporary analytic philosophy.

In the next chapter we will talk more about linguistic expressions and the things they denote (or refer to); but it is often confusing to the relative newcomers to philosophy to distinguish expressions and their denotations. For instance, all of the following are false:

- ‘Aristotle’ is a philosopher.
- Aristotle is a name.
- ‘Aristotle’ is the name of ‘Aristotle’.
- Aristotle is the name of Aristotle.

The correct things to say are as follows:

- Aristotle is a philosopher.
- ‘Aristotle’ is a name.
- ‘Aristotle’ is the name of Aristotle.

As you can see, possible confusion between words and things in the world is related to the use of quotation marks. So, to avoid later confusion, I would like to say a little about quotation to conclude this chapter. In particular, I would like to discuss three things: direct and indirect quotation, the use/mention distinction, and scare quotes. Along the way, I will also draw an important distinction between the object language and the metalanguage.

First, the kind of quotations that involve quotation marks are *direct quotations*. Generally, quotations are about reproducing words, but direct quotations reproduce the utterances or inscriptions of the original word by word. In contrast, indirect quotations do not usually involve quotation marks and convey the *meanings* of the original. For instance, compare:

<i>Direct quotation</i>	<i>Indirect quotation</i>
(a) Adam said, 'I will go to school today'.	Adam told me that he would go to school that day.
(b) Betty said, 'Did you go to school yesterday?'	Betty asked me if I had gone to school the day before.
(c) Charlie said, 'Come to school now!'	Charlie ordered me to go to school immediately.

An *indexical* (or an *indexical expression*) is an expression that denotes potentially different objects depending on the context, i.e., who uses the expression, when and where s/he uses it, etc. Pronouns such as 'I', 'you', 'he', 'she', and 'it',<sup>18</sup> adverbs such as 'today', 'yesterday', 'now', and 'here', and tensed verbs are all indexicals. As you can see above, indexicals in direct quotations need to be adjusted in the corresponding indirect quotations.

In the rest of this section, we will focus on direct quotations, setting aside indirect quotations. (We will come back to indirect quotation in Section 4.2.) Again, direct quotations involve quotation marks. What does a quoted expression denote? Consider:

- (d) Aristotle is a philosopher.
- (e) Aristotle was the teacher of Alexander the Great.
- (f) Aristotle is the most famous disciple of Plato.
- (g) 'Aristotle' is a name.
- (h) 'Aristotle' is a nine-letter word.
- (i) 'Aristotle' denotes Aristotle.

In (d) to (f), I am *using* the word 'Aristotle' to denote the Greek philosopher Aristotle, whereas in (g) to (i) (except the last token in (i)), I am *mentioning* the word 'Aristotle' to denote the word (type?) 'Aristotle'. That is, the used word 'Aristotle' is the name of the man Aristotle, whereas the mentioned word, the word in quotes, is the name of the word 'Aristotle'. Analogously,

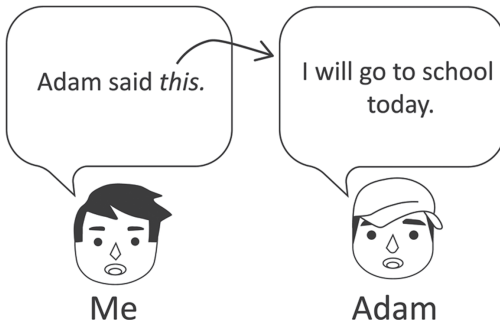


Figure 1.5 Adam said, 'I will go to school today'.

the quotations in (a) to (c) above denote the words (token?) that Adam, Betty, and Charlie uttered, respectively. For instance, the sentence I uttered in quotes in (a) plays the same role as '*this*' in Figure 1.5, which denotes the word token Adam produced.

Aristotle was indeed the teacher of Alexander the Great, so (e) is true. Since (f) is also true and Aristotle is the most famous disciple of Plato, we can replace 'Aristotle' in (e) with 'the most famous disciple of Plato' and retain the truth value of the sentence. But we cannot do the same with (h): 'Aristotle' is a nine-letter word. (h) is true, but if we replace 'Aristotle' with 'The most famous disciple of Plato', the resulting sentence, 'The most famous disciple of Plato' is a nine-letter word, is false. That's because in (h), those expressions are not the names of the man Aristotle but the names of different expressions in quotes. Similarly, if we replace 'I' in (a) with 'Adam', the resulting sentence will be false, for Adam did not literally say, 'Adam will go to school today'. Generally, in quotes we cannot substitute expressions that we can substitute when they are used.

In sum, while used expressions denote objects in the world, mentioned expressions, expressions in quotes, denote word types and tokens. This distinction is called *the use/mention distinction*.

By the way, what works as quotation marks? In writing, the single '...' (in British English) or double "... " (in American English) are typical quotation marks, but when quotes are frequently made, such as in books in linguistics, italics may be used in their stead (e.g., *Aristotle* instead of 'Aristotle'). This can be a source of confusion because in philosophy, italics are often used for other purposes. For instance, I may (and did) say: the predicate 'is a professor' stands for the property *being a professor*. In a context in which it is obvious that the relevant expression is quoted, the quotation marks may be omitted to avoid clutter. In fact, I did that a lot in Sections 1.3 and 1.4. This also includes the quotation of logical and mathematical symbolizations; we don't

usually put quotation marks around those symbolizations even when they are, strictly speaking, quoted.

It is generally more difficult in speech to make the existence of quotation marks explicit; but, if necessary, finger (or air) quotes (Figure 1.6) or the expression ‘quote ... unquote’ may be used. An excessive use of finger quotes, however, may make you an object of ridicule, so moderation is recommended.

Sometimes we use English but mention expressions in another language. This happens especially when we are talking about features of the other language. For example, I may say:

- Takeo said, ‘雪は白い (yuki wa shiroi)’.
- ‘雪 (yuki)’ is a noun and means *snow* in Japanese.
- ‘白い (shiroi)’ is an adjective and means *white* in Japanese.

In such cases, the language we are talking about, the language that is the object of our discussion, is called the *object language*, and the language we are using as the background language is called the *metalanguage*. (‘Meta’ means ‘beyond’ or ‘at a higher level’.)<sup>19</sup> Then, in the above example, Japanese is the object language and English is the metalanguage. Both object language and metalanguage can be any language; for example, the object language can be a formal language such as the language of logic, as we saw in Sections 1.3 and 1.4. When we are talking about English expressions in English, the object language and the metalanguage are identical, i.e., English.

Lastly, scare quotes. It was the British philosopher Elizabeth Anscombe who coined the term ‘scare quotes’ back in 1956. Frequent use of scare quotes is quite a contemporary phenomenon. Above I equated quoted words with mentioned words, but scare quotes are an exception to this equation: what’s inside scare quotes is used, not mentioned.

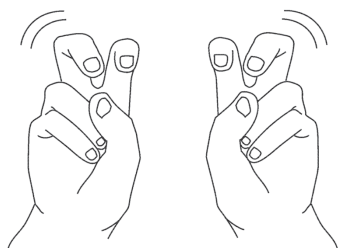


Figure 1.6 Finger quotes.

A good example is the following infamous allegation made on the internet by the US President Donald Trump against his predecessor:

- President Obama had my ‘wires tapped’ in Trump Tower.

Unlike the traditional quotations, here ‘wires tapped’ does not denote an expression. For one thing, you cannot replace it with *‘this’* and keep the sentence grammatical. For another thing, if you remove the quotation marks, the sentence will still make perfect sense.

For a few other examples involving scare quotes:

- Many of those ‘racists’ are in fact just trying to protect their investments.
- The college sent all the department chairs to the ‘self-awareness’ retreat.
- Leibniz once said that this world is ‘the best of all possible worlds’.
- Sarah boasts 562 Facebook ‘friends’.
- The ‘dead’ boy has been found alive.
- The Morning Star and the Evening Star – these ‘stars’ are actually one and the same *planet*, the planet Venus.
- My sister ‘taught’ me how to drive.

The role of scare quotes is to allow the writer/speaker to distance himself from the expression in the quotes for various reasons: he may mean it not literally but only figuratively or with irony or sarcasm, or he thinks that the expression is somehow inappropriate, inaccurate, or misleading. Scare quotes give the readers/audience the impression that the writer/speaker is quoting someone else’s words, even though that someone else may not exist. When it was shown that nobody was wiretapping Trump Tower, Donald Trump’s response was that he did not mean literally that somebody was wiretapping Trump Tower on Obama’s behalf; that’s why he put his assertion in scare quotes.

As you can see in this example, scare quotes often give you (bad) excuses when you get into trouble. Students who make too many scare quotes are perceived by their teachers and peers as lacking confidence in their own writing. Indeed, their writing tends to be bad. So, avoid scare quotes; find accurate expressions instead.

### Exercise Questions

(The exercise questions marked \* have answers at the end of the book.)

1. Explain the following concepts and distinctions.

Word token/word type; particulars/universals; properties/relations; realism/anti-realism; platonism/nominalism; validity; equivalence relation; numerical/relative identity; four-dimensionalism; necessary/sufficient condition; direct/indirect quotation; use/mention; object language/metalanguage; scare quotes.

2. Give examples of the following.
  - (a) Declarative sentence.
  - (b) Interrogative sentence.
  - (c) Imperative sentence.
  - (d) Exclamatory sentence.
  - (e) Proper name.
  - (f) Definite description.
  - (g) 1-place nominal predicate.
  - (h) 1-place adjectival predicate.
  - (i) 1-place verbal predicate.
  - (j) 2-place predicate.
  - (k) 3-place predicate.
  - (l) 0-place predicate.
- \*3. Translate the following English sentences into the language of propositional logic. Keys: *L*: My dog loves me; *H*: I am happy; *S*: I am sad.
  - (a) My dog loves me and I am happy.
  - (b) My dog loves me or I am sad.
  - (c) I am happy if and only if my dog loves me.
  - (d) If I am not happy, then I am sad.

Translate the following symbolizations into English and determine their truth values, assuming that all atomic sentences, *L*, *H*, and *S*, are true.

  - (e)  $S \rightarrow \neg L$
  - (f)  $(H \vee S) \rightarrow L$
  - (g)  $(L \rightarrow H) \rightarrow (L \vee S)$
  - (h)  $(H \wedge S) \rightarrow (H \wedge \neg H)$
- \*4. Draw truth tables for the following sentences.
  - (a)  $\neg\neg P$
  - (b)  $P \leftrightarrow \neg\neg P$
  - (c)  $P \wedge \neg P$
  - (d)  $P \vee \neg P$
  - (e)  $\neg P \wedge Q$
  - (f)  $\neg(P \wedge Q)$
  - (g)  $(P \wedge Q) \leftrightarrow (P \rightarrow \neg Q)$

\*5. Translate the following into the language of predicate logic. Keys.

$a$ : Adam;  $b$ : Buster;  $c$ : Chris;  $Tx$ :  $x$  is tall;  $Sxy$ :  $x$  is shorter than  $y$ .

- (a) Adam is tall, and Buster isn't tall.
- (b) If Buster is tall, then Buster isn't shorter than Adam.
- (c) Adam, Buster, or Chris is tall.
- (d) Chris is not shorter than Adam or Buster.
- (e) Chris is shorter than Adam only if he is shorter than Buster.  
(If you are not sure how 'only if' works, see note 5.)

Translate the following symbolizations into English and determine their truth values, assuming that Adam is tall, but Buster and Chris are not; Chris is shorter than Buster, and Buster is shorter than Adam.

- (f)  $Sbc \vee Scb$
- (g)  $Tc \rightarrow Sbc$
- (h)  $\neg(Sab \wedge Scb)$
- (i)  $(Ta \vee Tb) \wedge (Tb \vee Tc)$
- (j)  $(Tb \vee Scb) \rightarrow Tb$

\*6. Suppose: if I go to work, I will have money; and I am happy only if I have money.

- (a) Which of the following are a necessary condition, a sufficient condition, a necessary and sufficient condition, or neither, for 'I go to work'? (A) I have money. (B) I am happy.
- (b) Same question for 'I have money'. (A) I go to work. (B) I am happy.
- (c) Same question for 'I am happy'. (A) I have money. (B) I am happy.

\*7. Put quotation marks "..." (double) where necessary to make the following sentences true (or as true as possible). If the sentence does not need any quotation marks, state 'Unnecessary' clearly. If you need quotation marks inside a quote "...," use single quotation marks '...'.  
(a) The most famous disciple of Plato is Aristotle.

- (b) The most famous disciple of Plato is identical with Aristotle.
- (c) The most famous disciple of Plato refers to Aristotle.
- (d) The most famous disciple of Plato refers to the same thing as Aristotle does.
- (e) The most famous disciple of Plato and Aristotle are one and the same man.
- (f) The denotation of the most famous disciple of Plato is Aristotle.
- (g) John said, the most famous disciple of Plato is Aristotle.
- (h) John said, the most famous disciple of Plato denotes Aristotle.

- (i) John told Ann that the most famous disciple of Plato denoted Aristotle.
- (j) Being tall and being a dog are both properties.
- (k) Is tall and is a dog are both predicates.
- (l) The most famous disciple of Plato is Aristotle is true.
- (m) It is true that the most famous disciple of Plato is Aristotle.
- (n) The post office has thousands of letters sent every day.
- (o) The post office has only 12 letters.
- (p) Adam knows Karate.
- (q) Adam knows Karate and three other Japanese words.

\*8. Explain what the Knight is saying below:

“You are sad,” the Knight said in an anxious tone: “let me sing you a song to comfort you.”

“Is it very long?” Alice asked, for she had heard a good deal of poetry that day.

“It’s long,” said the Knight, “but it’s very, *very* beautiful. Everybody that hears me sing it – either it brings the *tears* into their eyes, or else –”

“Or else what?” said Alice, for the Knight had made a sudden pause.

“Or else it doesn’t, you know. The name of the song is called ‘*Haddocks’ Eyes*’”.

“Oh, that’s the name of the song, is it?” Alice said, trying to feel interested.

“No, you don’t understand,” the Knight said, looking a little vexed. “That’s what the name is *called*. The name really is ‘*The Aged Man*’”.

“Then I ought to have said ‘That’s what the *song* is called?’” Alice corrected herself.

“No, you oughtn’t: that’s quite another thing! The *song* is called ‘*Ways and Means*’: but that’s only what it’s *called*, you know!”

“Well, what *is* the song, then?” said Alice, who was by this time completely bewildered.

“I was coming to that,” the Knight said. “The song really is ‘*A-sitting On A Gate*’: and the tune’s my own invention.”

So saying, he stopped his horse and let the reins fall on its neck: then, slowly beating time with one hand, and with a faint smile lighting up his gentle foolish face, as if he enjoyed the music of his song, he began. (Lewis Carroll, *Through the Looking-Glass*, Chapter VIII, 1871)



## Suggested Further Reading

In the subsequent Suggested Further Reading sections, I will suggest only readings that are appropriate for the targeted readers of this book to read as a next step. I will not suggest readings that are either too basic or too advanced for those readers.

In general, the following encyclopedias will be very helpful:

- *Stanford Encyclopedia of Philosophy* (<https://plato.stanford.edu>).
- *Internet Encyclopedia of Philosophy* ([www.iep.utm.edu](http://www.iep.utm.edu)).

These online encyclopedias have numerous articles that cover all areas of philosophy. The articles are usually well-balanced and thorough, so if you are interested in any topic in philosophy, you cannot go wrong by reading some of the articles related to it. At the same time, however, most of the articles are written for graduate students and professional researchers and not for undergraduate students or those who are the targeted readers of this book. The latter may find many of the articles too long, too detailed, and too complicated. If you do find an article too much to handle, move on to something else, likely to be mentioned in the article.

*Wikipedia* articles in philosophy are also often helpful. They are generally less authoritative than articles in the above encyclopedias and may contain more errors, but they are usually much more accessible and can jumpstart your research.

In connection to this chapter,

- D. H. Mellor and Alex Oliver (eds.), *Properties*

contains influential articles on universals and properties. (See Bibliography for more bibliographical information.)

- Alyssa Ney, *Metaphysics*

is a contemporary introduction to metaphysics for undergraduate students. It contains fuller discussions about many of the metaphysical issues touched on in this and other chapters of this book.

If you want to own (and study) an introductory logic textbook,

- John Nolt, Dennis Rohatyn, and Achille Varzi, *Logic*, 2nd edn,

is a good inexpensive(!) textbook for both propositional and predicate logic.

- John Hawthorne, “Identity”

further explores the issues pertaining to identity.

## Notes

- 1 In this book, I will set aside the possibility of abstract particulars of a certain kind, called *tropes*. The idea of tropes was first introduced by D. C. Williams (1953a, 1953b). Mathematical objects, such as sets and numbers, are also abstract objects, as they exist outside of spacetime. I will touch on mathematical objects in several places in this book (most conspicuously in Sections 2.2 and 2.6), but will have to be a little vague about their nature and their relation with the particular/universal distinction.
- 2 Universals understood this way are called *transcendent* universals. In contrast, on the conception of *immanent* universals, a universal is located in spacetime, located where its instances (i.e., particulars) are located, and multiply located if its instances are located at different places (Armstrong 1978, Lewis 1986). I think that when most people think about universals, they embrace the transcendent conception; so I focus on that conception in this book, setting aside the immanent conception.
- 3 However, contemporary platonists are usually not committed to the idea that universals are perfect whereas particulars are imperfect.
- 4 ‘Or’, used this way, is called an *inclusive* ‘or’:  $P$  or  $Q$ , including the case in which  $P$  and  $Q$  are both true. So, on this usage, ‘ $P$  or  $Q$ ’ is true if  $P$  is true and  $Q$  is true. In contrast, an *exclusive* ‘or’ is such that if  $P$  is true and  $Q$  is true, ‘ $P$  or  $Q$ ’ is false. Suppose Adam had both soup and salad before the main meal. Then ‘Adam had soup or salad’ is (doubly) true if the ‘or’ involved is an inclusive ‘or’, but false if the ‘or’ is exclusive. ‘Or’ used in logic is inclusive.
- 5 Unfortunately, logical notations have not been standardized and vary from textbook to textbook.  $\text{Not } P$  can be written as  $\sim P$  or  $\neg P$ .  $P$  and  $Q$  can be written as  $P \& Q$  or  $P \cdot Q$ .  $\text{If } P \text{ then } Q$  and  $P \text{ iff } Q$  can be written as  $P \supset Q$  and  $P \equiv Q$ , respectively. Throughout this book, ‘iff’ is an abbreviation of ‘if and only if’. ‘Only if’ is the opposite of ‘if’; so  $P \text{ only if } Q$  is equivalent to  $Q \text{ if } P$ , which is equivalent to  $\text{if } P \text{ then } Q$ .
- 6  $\forall$  (the upside-down A) comes from ‘for All  $x$ s, ...’, and  $\exists$  (the mirror image of E) comes from ‘there Exists some  $x$  such that ...’. Some textbooks use  $(x)$  instead of  $\forall x$  for the universal quantifier. If they do, then the matching existential quantifier should be  $(\exists x)$  (with parentheses).
- 7 Generalized De Morgan’s Laws are the generalizations of the original De Morgan’s Laws in propositional logic. In many respects, the universal quantifier works like a generalized conjunction, and the existential quantifier works like a generalized disjunction; for instance,  $\forall xTx$  (everything is tall) means  $Ta \wedge Tb \wedge Tc \wedge \dots$  (Adam is tall and Betty is tall and Charlie is tall and ...), and  $\exists xPx$  (something is a professor) means  $Pa \vee Pb \vee Pc \vee \dots$  (Adam is a professor or Betty is a professor or Charlie is a professor or ...). Because of that, a few textbooks use  $\wedge x$  and  $\vee x$  instead of  $\forall x$  and  $\exists x$  for the universal and the existential quantifier.
- 8 Bertrand Russell (1905) famously objected to the idea that definite descriptions are singular terms. His own theory is called *Russell’s theory of definite descriptions*.

We will set aside the theory in this book because one has to have a firm grasp of predicate logic to fully appreciate it. Many logic textbooks contain an account of Russell's theory. Russell's original paper (1905) is rather difficult to read and may be avoided.

Some proper names, such as 'the United States of America', 'the United Kingdom', 'the Holy Roman Empire', 'the Fountain of Youth', 'the Morning Star', and 'the Evening Star' (both names of Venus), contain 'the' and expressions not literally descriptive; thus, for instance, the United States of America and the United Kingdom do not need to be united to be so called. The Morning Star and the Evening Star are not stars but a planet. We will ignore any descriptive element which might be left over in those proper names and treat them as totally non-descriptive.

- 9 There are also cases in which 'a(n) + such and such' is used as a quantificational phrase, meaning basically the same thing as 'every + such and such' or 'some (or one) + such and such'. For instance, 'a podiatrist' in 'A podiatrist is a medical doctor' is an instance of the former (but not 'a medical doctor', which is part of the predicate 'x is a medical doctor'), and 'A podiatrist in that hospital was instrumental to solving my foot problem' is an instance of the latter.
- 10 The formal meanings of 'denote' and 'denotation' will be given in Chapter 2. In this chapter I use those words informally, counting on your natural understanding of the words.
- 11 Whether every (possible) predicate denotes a property or not is debatable. For instance, the predicates 'x is a dog' and 'x is a table' may denote properties, but how about the predicates 'x is not a dog' and 'x is a dog or a table'? An *abundant theory of properties* generally holds that there is an abundance of properties in the world; so, possibly, there is a property for each predicate. A *sparse theory of properties*, in contrast, posits relatively few properties in the world. For instance, Armstrong (1978, 1989) maintained that the only properties that exist are 'natural' properties uncovered by science. In a sparse theory like Armstrong's, it is unlikely that every predicate denotes a property. We will not discuss various theories of properties in this book; but, for the sake of simplicity, I assume properties are abundant. I will say a little more about properties in Section 3.7 (essential vs accidental properties), Section 5.8 (intrinsic vs extrinsic properties), and Section 5.9 (mental vs physical properties).
- 12 Whether *abstract nouns (and noun phrases)*, such as 'beauty', 'extreme beauty', 'dogness', 'catness', and 'humanity', denote particulars or universals is actually a thorny question, although I have been speaking as if they denote universals. Some philosophers and linguists think they denote universals; in particular, the predicate 'x is a dog' and the abstract noun 'dogness' denote one and the same universal dogness. Other philosophers and linguists think abstract nouns actually denote particulars that correspond to universals, things that descended from the platonistic heaven, as it were; so 'dogness' denotes some particular that corresponds to the universal 'x is a dog' denotes. (Presumably, this particular still does not have a spatiotemporal location.) Confusing, isn't it? So I will set aside this issue in this book.

- 13 By the way, quite a few students say and even write ‘one *in* the same’. (True, ‘one and the same’ sounds like ‘one in the same’.) But the last time I checked, it was still considered a mistaken usage; so you’d better avoid it. (But who knows what will happen in a few years from now? The current usage of ‘begging the question’, meaning ‘inviting the question’, was considered a mistaken usage only a couple of decades ago. The original meaning, as in ‘a question-begging argument’, was ‘assuming what ought to be proved’.)
- 14 The symbolization of this statement looks like this:

$$\forall x \forall y (x = y \leftrightarrow \forall X (Xx \leftrightarrow Xy)).$$

You probably have never seen the symbol  $\forall X$  (with an uppercase  $X$ ) before. This sentence is a sentence of *second-order logic*. Recall that predicate (or quantificational) logic, which I explained in Section 1.4, is also called *first-order logic*. Quantifiers in first-order logic such as  $\forall x$  quantify over individuals (or particulars): ‘for any  $x$ ’ means ‘for any individual  $x$ ’. In contrast, second-order quantifiers such as  $\forall X$  quantify over properties: ‘for any  $X$ ’ means ‘for any property  $X$ ’. Second-order logic, or, more generally, *higher-order logic*, is too complicated to be dealt with in this book. But it is important to keep in mind that there are English sentences, like Leibniz’s Law, that can be expressed only in the language of second-order logic.

- 15 Sometimes only the former is called ‘Leibniz’s Law’.
- 16 Quod erat demonstrandum (Latin for ‘that which was to be demonstrated’). Often added at the end of a proof in mathematics.

Given that the Identity of Indiscernibles is derivable from the basic logical laws constitutive of the concept of identity, you may find it puzzling that quite a few philosophers question the validity of the law and even try to give counterexamples (Black 1952). Consciously or unconsciously, those philosophers must be excluding *being identical with  $x$*  and its kin from the relevant set of properties. Recall sparse theories of properties mentioned in Note 11.

- 17 There is a complication, however. Suppose there is indeed a ship both the Original Ship (existing only until T1 by definition) and the Repaired Ship (existing only after T2 by definition) are parts of; name it S1. Suppose also that there is another ship both the Original Ship and the Reassembled Ship (existing only after T2 by definition) are parts of; name it S2. Stipulate that Theseus’s Ship (to be distinguished from the Original Ship) is the ship that persists for a period of time including T1 and T2. Is Theseus’s Ship, then, identical with S1 or S2? This question is a question of numerical identity.

There are many so-called ‘identity puzzles’ in philosophy, and many of them are confusing in large part because, when they are presented, the relevant terms (such as ‘Theseus’s ship’, ‘the original ship’, ‘the repaired ship’, etc.) are not clearly defined. (I defined my terms very clearly – I hope.) When you encounter those puzzles, make sure how the relevant terms are (or are not) defined.

- 18 If these pronouns are used when you are directly pointing to objects, they are called ‘demonstratives’ (Section 1.4).

- 19 Because of this, people tend to think that metaphysics is so-called because it is the study of higher-level beings that physics doesn't deal with. However, the real origin of the word 'metaphysics' is more mundane: a Roman editor who assembled pieces of Aristotle's work on nature into the book called *Physics* assembled its leftovers into the next book, which was called *Metaphysics* (= 'after *Physics*').

# Extension and Intension

---

## 2.1 Introduction

In Chapter 1 we have decided to assume, at least for the sake of discussion, that abstract objects exist, i.e., that there are things that exist outside of spacetime. The next question, then, is: What are abstract objects? What are they like? In this chapter, I will present two kinds of abstract objects: *sets* and *intensional objects*. (Note the spelling of ‘intensional’; it’s not ‘intentional’ but ‘intensional’ with an ‘s’.) Along the way, I will introduce you to three mathematical/philosophical theories: set theory, mereology, and possible worlds semantics.

I will first tell you an interesting fact, made famous by the American philosopher Willard V. Quine, and then ask you a question. A *cordate*,<sup>1</sup> by definition, is an animal with a heart. A *renate* is an animal with a kidney. An interesting fact is: all cordates happen to be renates, and all renates happen to be cordates in reality; that is, every animal which has a heart also has a kidney, and every animal which has a kidney also has a heart, even though there is no logical or biological reason why that ought to be so. Then, do the predicates ‘*x* is a cordate’ and ‘*x* is a renate’ stand for one and the same abstract object, or do they stand for two different abstract objects?

In my experience, the majority of students would answer, ‘Two distinct abstract objects’, although there are always a few who would say, ‘One and the same abstract object’. But here, just as in Section 1.1 with the number of animals in my backyard, who is right and who is wrong is not as important as the question: How do you describe the difference? The answer is: ‘*x* is a cordate’ and ‘*x* is a renate’ *denote* the same *set* but *connote* different *intensional objects*. But what is a set, and what is an intensional object? I will answer the first question first.

## 2.2 Set Theory

Set theory, first developed by the German mathematician Georg Cantor and formalized in the first half of the twentieth century, lies at the foundation of mathematics along with logic. The details of the theory are rather

complicated, but you do not need to know those details in order to do philosophy (unless you do philosophy of mathematics). The core part of set theory you need to know for philosophical purposes is pretty simple.

A *set* is a group of things, *any* things, including other sets. The single most basic relation in set theory, by which all the other set-theoretic properties and relations are defined, is the *set-membership* relation: object  $a$  is a member of set  $S$  (or  $a$  is in  $S$ ), symbolized as  $a \in S$  and illustrated as follows (Figure 2.1).

Its denial is that  $a$  is not a member of  $S$  (or  $a$  is not in  $S$ ),  $a \notin S$  (Figure 2.2).

A set may be written like  $\{a, b, c \dots\}$ ; this means that  $a, b, c \dots$  are (all) the members of the set. Also, a set may be written as, e.g.,  $\{x: x \text{ is a table}\}$ ; this means the set of (all) tables. (The word ‘all’ is often omitted in this context; so ‘the set of  $x$ s’ mean ‘the set of *all*  $x$ s’.) Sets are abstract objects and thus are not located anywhere in spacetime.

The set that has no member, i.e.,  $\{\}$ , exists and is called *the empty (or null) set*, symbolized as  $\Phi$ . For instance, the set of square circles is the empty set, for there is no such thing as a square circle. A set that has only one member, e.g.,  $\{a\}$ , i.e., the set containing the single object  $a$ , is called a *unit set (or singleton)*.  $\{a\}$  is an abstract object different from the individual  $a$ . For instance, Tibbles may be sitting at the center of my yard, but  $\{\text{Tibbles}\}$ , the set of cats in my yard, does not exist anywhere in spacetime.

Here is one of the most important principles in set theory:

- The Axiom of Extensionality  
Set  $P$  and set  $Q$  are different sets iff at least some of their members are different.

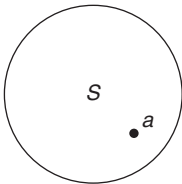


Figure 2.1  $a \in S$ .

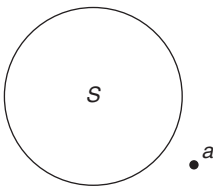


Figure 2.2  $a \notin S$ .

Put the other way around,  $P$  and  $Q$  are one and the same set iff their members are identical. This means that the identity of a set is determined solely by its members and nothing else. So the set of cordates  $\{x: x \text{ is a cordate}\}$  and the set of renates  $\{x: x \text{ is a renate}\}$  are one and the same set because their members are identical. This is what we mean when we say *that sets are extensional objects*. The set of cordates (= the set of renates) is *the extension* of the predicate 'x is a cordate' (or 'x is a renate'). An extension may be thought of as a group of things enclosed by an imaginary extended picket fence.<sup>2</sup>

A set, say  $P$ , can itself be a member of another set, say  $Q$ :  $P \in Q$ . The set-membership, however, is generally not transitive: generally,  $x \in P$  and  $P \in Q$  do not mean  $x \in Q$ . For instance, you may be a member of the Philosophy Club, and the Philosophy Club may be a member of the Association of the University Clubs, but you are not a member of the Association of the University Clubs.

Finally, I will define some important set-theoretic concepts that are definable in terms of set-membership. The *complement* of set  $P$ ,  $-P$  (or  $\bar{P}$ ), with respect to the domain  $D$  is  $D - P$ , where the *domain* is the set of all things under consideration (Figure 2.3). So for any  $x \in D$ ,  $x \in -P$  iff  $x \notin P$ ; or  $-P = \{x \in D: x \notin P\}$ .

If  $P$  is the set of professors and  $D$  is the set of all people,  $-P$  is the set of all people who are not professors.

The *intersection* of sets  $P$  and  $Q$ ,  $P \cap Q$ , is the set corresponding to the overlapping area of  $P$  and  $Q$  (Figure 2.4). So, for any  $x$ ,  $x \in P \cap Q$  iff  $x \in P$  and  $x \in Q$ ; or  $P \cap Q = \{x: x \in P \text{ and } x \in Q\}$ . The intersection of  $P$  and  $Q$  is also called the *product*, or *meet*, of  $P$  and  $Q$ .

If  $P$  is the set of professors and  $Q$  is the set of Quakers,  $P \cap Q$  is the set of Quaker professors.

The *union* of sets  $P$  and  $Q$ ,  $P \cup Q$ , is the set covering both  $P$  and  $Q$  (Figure 2.5). So, for any  $x$ ,  $x \in P \cup Q$  iff  $x \in P$  or  $x \in Q$ ; or  $P \cup Q = \{x: x \in P \text{ or } x \in Q\}$ . When the intersection of  $P$  and  $Q$  is called the *product* of  $P$  and  $Q$ , the union of  $P$  and  $Q$  is called the *sum* of  $P$  and  $Q$ ; when the intersection is called the *meet* of  $P$  and  $Q$ , the union of  $P$  and  $Q$  is called the *join* of  $P$  and  $Q$ .

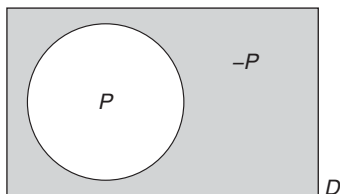
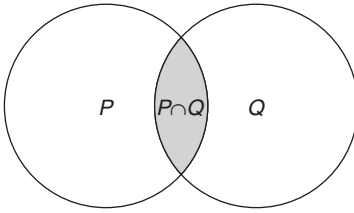
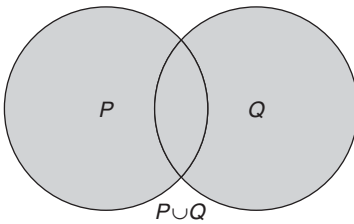
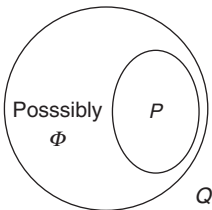


Figure 2.3  $-P$ .



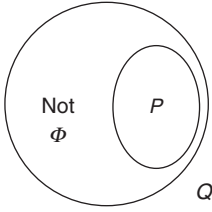
Figure 2.4  $P \cap Q$ .Figure 2.5  $P \cup Q$ .Figure 2.6  $P \subseteq Q$ .

If, again,  $P$  is the set of professors and  $Q$  is the set of Quakers,  $P \cup Q$  is the set of people who are either professors or Quakers.

Set  $P$  is a *subset* of set  $Q$ , i.e.,  $P \subseteq Q$ , iff  $P$  is the same set as  $Q$  or is totally contained in  $Q$  (Figure 2.6). So,  $P \subseteq Q$  iff for any  $x$ , if  $x \in P$  then  $x \in Q$ .

Set  $P$  is a *proper subset* of set  $Q$ , i.e.,  $P \subset Q$ , iff  $P$  is a subset of  $Q$  but not the same set as  $Q$  (Figure 2.7). In other words,  $P \subseteq Q$  iff  $P \subset Q$  or  $P = Q$ .

So every set is a subset of itself, but not a proper subset. (We could say that every set is an (or the) *improper* subset of itself.) Suppose all professors are Quakers but not all Quakers are professors; then the set of professors is a proper subset of the set of Quakers.

Figure 2.7  $P \subset Q$ .

As I said, set theory and logic are the two foundations of all mathematics, and they are also closely related to each other. The set-theoretic concepts I have just defined except that of proper subset, i.e., the concepts of complement, intersection, union, and subset, correspond to negation, conjunction, disjunction, and conditional in logic in the following way: Suppose that the extension of the predicate  $Px$  (say, ' $x$  is a professor') is the set  $P$  of all professors while the extension of the predicate  $Qx$  (say, ' $x$  is a Quaker') is the set  $Q$  of all Quakers. Then the extension of  $\neg Px$  (' $x$  is not a professor') will be the complement of  $P$ ,  $\neg P$ ; the extension of  $Px \wedge Qx$  (' $x$  is a Quaker professor') will be the intersection of  $P$  and  $Q$ ,  $P \cap Q$ ; the extension of  $Px \vee Qx$  (' $x$  is either a professor or a Quaker') will be the union of  $P$  and  $Q$ ,  $P \cup Q$ ; and the extension of  $Px \rightarrow Qx$  ('if  $x$  is a professor, then  $x$  is a Quaker'), which is equivalent to  $\neg Px \vee Qx$ , will be the union of  $Q$  and the complement of  $P$ ,  $\neg P \cup Q$ . This set is the domain  $D$  itself (i.e., everything in the domain satisfies  $Px \rightarrow Qx$ ) when and only when  $P$  is a subset of  $Q$ , i.e.,  $P \subseteq Q$ . To summarize:

- $\neg P$  corresponds to  $\neg Px$ ;
- $P \cap Q$  corresponds to  $Px \wedge Qx$ ;
- $P \cup Q$  corresponds to  $Px \vee Qx$ ;
- $P \subseteq Q$  corresponds to  $Px \rightarrow Qx$ .

Note that the way  $P \subseteq Q$  corresponds to  $Px \rightarrow Qx$  is a little different from the way  $\neg P$ ,  $P \cap Q$ , and  $P \cup Q$  correspond to  $\neg Px$ ,  $Px \wedge Qx$ , and  $Px \vee Qx$ , respectively. This is because  $\neg P$ ,  $P \cap Q$ , and  $P \cup Q$  are themselves sets, whereas  $P \subseteq Q$  is merely a relation between the sets  $P$  and  $Q$ .

In sum, sets are abstract but extensional objects. Their identity is completely determined by their members and nothing else. Thus, the set of cordates and the set of renates are one and the same set.

### 2.3 Mereology

Sometimes the concept of set-membership is confused with another concept, that of *part-whole (or part-of) relation*. So, to make the concept of set

perfectly clear, I would like to talk, albeit briefly, about part-whole relations before moving on to intensional objects. The study and theory of part-whole relation is called *mereology*, where ‘mere-’ means ‘part’ in Greek. Mereology is itself a subject of much debate in contemporary philosophy. So it will be good for you to have some idea about the theory. Our discussion will focus on the difference between set-membership and the part-whole relation, and between sets and so-called *mereological sums* (or *fusions*).

While sets can be sets of anything, including other sets, mereology is mostly<sup>3</sup> concerned with spatiotemporal objects, objects that exist in spacetime. So let’s think about my own body. My body is made of body parts. For the sake of simplicity, let’s assume that my body is made of six body parts: one head, one torso, the right and the left arm, and the right and the left leg. (I ignore the fact that I may have the fourth dimension, the temporal dimension, according to four-dimensionalism, as we discussed in Section 1.5.) Then the set of my body parts has six members and can be depicted as follows (Figure 2.8).

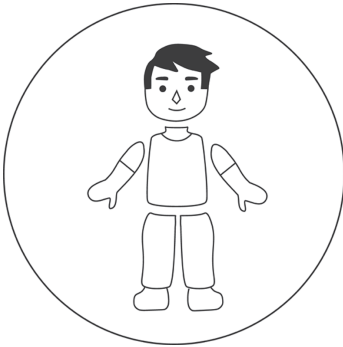


Figure 2.8 The set of my body parts.

Since the identity of sets is determined solely by its members, this is the same set (Figure 2.9).



Figure 2.9 The same set of my body parts.

This set does not have any spatiotemporal location even though its members do.

In comparison, my whole body is a material object and exists in a specific spatiotemporal region (Figure 2.10).

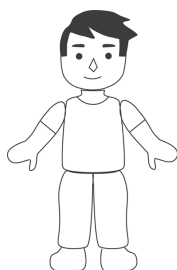


Figure 2.10 My body as a sum of my body parts.

The six body parts of my body are *mereological parts* of the body (the whole); conversely, the body is a *mereological sum* (or *fusion*) of the six parts. Unlike the set of the six parts we considered above, this mereological sum of the six parts is itself a concrete spatiotemporal object. Any spatiotemporal region occupied by one of the parts is also occupied by the whole body, and any spatiotemporal region occupied by the whole body is also occupied by one of the parts. When I say that our solar system is a part of the Milky Way galaxy, or that Earth is a part of our solar system, I am talking about mereological parts. (In contrast, when I say that Earth is a member of the set of planets in our solar system, I am talking about the set-membership.)

Parts can overlap. For instance, the whole consisting of my head, torso, right arm, and right leg (i.e., my right-side body) and the whole consisting of my head, torso, left arm, and left leg (my left-side body) are both parts of my whole body, partially overlapping each other, sharing my head and torso. The mereological sum of those two parts is my whole body. My whole body is also a part of my whole body, an *improper* part. In contrast, the aforementioned six parts of my body are *proper* parts of my body. This proper/improper distinction is analogous to that in subsets.

Compare this whole with the next whole (Figure 2.11). Here the six body parts of mine are cut off and reattached in a Frankensteinian fashion. This whole is a different mereological sum even though it is also a sum of the same six parts. Among other things, it occupies a different spacetime region.

In sum, while the set of the six parts does not exist in spacetime, their mereological sums do, and how they are connected or located with respect to one another makes different mereological sums.

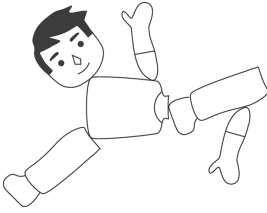


Figure 2.11 A different sum of my body parts.

Furthermore, unlike set-membership, the part-whole (or part-of) relation is transitive: if  $a$  is a part of  $b$  and  $b$  is a part of  $c$ , then  $a$  is a part of  $c$ . For instance, my eyes, nose, mouth, and ears are parts of my head, and my head is a part of my body; it follows that my eyes, nose, mouth, and ears are also parts of my body. Since Earth is a part of our solar system, and our solar system is a part of the Milky Way galaxy, Earth is a part of the Milky Way galaxy.

To repeat, sets of spatiotemporal objects are themselves not spatiotemporal objects but abstract objects; in contrast, mereological sums of spatiotemporal objects are themselves spatiotemporal objects. Because of this, some nominalists about abstract objects have tried to replace set theory with mereology as a foundation of mathematics. Most mathematicians now agree, however, that mereology is no substitute for set theory. We need abstract objects in mathematics. Still mereology receives much attention in contemporary philosophy because it is related to the issues of *material constitution*, i.e., how smaller material parts can make up larger wholes.

Two spatiotemporal objects need not be spatiotemporally conjoined (or adjacent) to each other to create their mereological sum. Two spatiotemporally disjoint gloves (shoes, socks, etc.) can create one spatiotemporal object, a pair of gloves (shoes, socks, etc.). Even two seemingly spatiotemporally conjoined objects, such as my head and my torso, are not really touching each other at the microscopic level. There is so much empty space between two molecules, two atoms, or two subatomic particles. If two objects need not be spatiotemporally conjoined to compose their sum, what are the restrictions of composition? Here is one proposed answer:

- The Principle of Unrestricted Mereological Composition  
There is no restriction on mereological composition; *any* group of spatiotemporal objects composes their mereological sum.

So even a group of spatiotemporally very disparate objects, such as my body, the Eiffel Tower, and one particular photon coming out of the sun, have their

sum, according to the Principle. The theory of mereological composition that embraces the Principle is called *mereological universalism*.<sup>4</sup> Whether this Principle is correct or not, and, if not, what the proper restrictions on mereological composition ought to be, are two of the major questions in philosophical mereology.

## 2.4 Kinds of Extension and Intension

That's all for mereology. Going back to our original question raised in Section 2.1, it is now clear in what sense 'x is a cordate' and 'x is a renate' stand for the same abstract object: they stand for the same *set*. Many people, however, would contend that they stand for different abstract objects. They would say something like this: We have agreed earlier that the 1-place predicates 'x is a cordate' and 'x is a renate' stand for *properties*. But the property *being a cordate* and the property *being a renate* are different properties. The first is the property *being an animal with a heart* and the second is the property *being an animal with a kidney*. It is just a pure coincidence that all and only creatures that have hearts have kidneys; so the predicates 'x is a cordate' and 'x is a renate' in fact instantiate two different properties.

Properties, understood this way, distinguished from sets, are *intensional objects* of sort. Just as sets are not the *only* extensional objects, properties are not the *only* intensional objects; there are some other extensional and intensional objects, as you will see shortly. 'Intensional' is the antonym (the opposite word) of 'extensional'. Generally, something more is *involved* in intension than the corresponding extension; so intensional objects, such as the properties *being a cordate* and *being a renate*, are more fine-grained objects than the corresponding extensional objects, such as the set of cordates/renates.

Again, properties are not the only intensional objects. I've said that when we talk about linguistic expressions, we will talk about three kinds: singular terms, (1-place) predicates, and (declarative) sentences as combinations of singular terms and (1-place) predicates. (I've also said that I will often omit the adjectives '1-place' and 'declarative'.) Just as predicates have sets as extensions and properties as intensions, singular terms and sentences also have extensions and intensions.

But first, extension and intension have other names. Unfortunately, the terminology in this area is not standardized and is rather complicated. Generally, extensions of linguistic expressions may be called their 'denotations', and when they are, the corresponding intensions are called 'connotations'. We also use the verbs 'denote' and 'connote' correspondingly. So the predicates 'x is a cordate' and 'x is a renate' denote the same set (or extension) but connote different properties; their denotations are one and the same but their connotations are different. This is consistent with our ordinary usage of

the word 'connote'. 'Cordate' and 'renate' have different connotations, one something to do with hearts and the other something to do with kidneys. 'Spinster' and 'bachelorette' denote the same set of unmarried adult females, but their connotations are quite different, aren't they? (Spinsters spin threads to make their own living.)

Gottlob Frege (1892), who first made this important distinction, called denotations 'references' (or 'Bedeutung' in German) and connotations 'senses' (or 'Sinn'); so, on this usage, a reference is an object denoted by an linguistic expression, and not an act of denoting. He also used the verb 'refer to' ('bedeuten') instead of 'denote' – the usage many contemporary philosophers embrace. According to Frege, reference and sense are the two elements of linguistic meaning, and every meaningful expression has a sense (it 'makes sense'), but some meaningful expressions do not have a reference. For instance, 'Pegasus', 'the Fountain of Youth', and 'the largest prime number' do have senses but not references because there is no such thing as Pegasus, the Fountain of Youth, or the largest prime number. These ideas of Frege's have been carried over to the contemporary distinction between extension and intension. Some people call the intension of an expression simply the 'meaning' of the expression.

If two expressions (e.g., predicates, singular terms, and sentences) have the same extension/denotation/reference, we say that they are co-extensional or co-referential; if two expressions have the same intension/connotation/sense, we say that they are co-intensional or *synonymous* (= meaning the same).

The words 'extension' and 'denotation' (or 'reference') can be used more narrowly. In the narrow sense, an extension is an extension only of a predicate, in contrast to extensions of singular terms and sentences. (Again, imagine a group of things enclosed by an extended picket fence.) 'Denotation' (or 'reference') can be used for denotations (references) only of singular terms. In what follows, however, we will usually use these terms more broadly and generally.

The extension (denotation, or reference) of a singular term is the individual object the term denotes while its intension is what I call an *individual concept* (after Carnap 1947), i.e., what the individual in question is conceived as. For instance, the singular terms 'Aristotle', 'the teacher of Alexander the Great', 'the most famous disciple of Plato', 'the author of *Metaphysics*', and 'the most influential philosopher in antiquity', are co-denotational, denoting one and the same individual, the philosopher Aristotle, but they have different connotations and express different individual concepts; setting aside the proper name 'Aristotle' for a moment, the second relates him with Alexander the Great, the third with Plato, the fourth describes the individual as an author, and the fifth as an influential philosopher.

The idea that singular terms have senses as well as references was convincingly argued for by Frege. Compare the identity statements ' $a = a$ '

with ' $a = b$ ' when  $a$  is indeed identical with  $b$ . For instance, compare 'The teacher of Alexander the Great is (identical with) the teacher of Alexander the Great' with 'The teacher of Alexander the Great is (identical with) the most famous disciple of Plato'. If the references of those singular terms were all the terms mean, then there should be no difference in the meanings of the singular terms; so there should be no difference in the meanings of those identity statements. But in fact there is: the former is a so-called logical truth whose truth can be known a priori (i.e., independently of any experience), while the latter is a non-logical truth whose truth can be known only a posteriori (i.e., through some experience). (For the meanings of 'logical', 'a priori', and 'a posteriori', see Chapter 3.) The difference in the meanings of 'the teacher of Alexander the Great' and 'the most famous disciple of Plato' is in the difference in their senses.

It is unclear what sense, what individual concept, the proper name 'Aristotle' expresses, as opposed to definite descriptions such as 'the teacher of Alexander the Great' and 'the most famous disciple of Plato'. Frege, however, maintained that, vague as they may be, proper names also must have senses. The above argument of his for the senses of definite descriptions is applicable also to proper names. Recall the example of Hesperus and Phosphorus, given in Section 1.5. 'Hesperus' is the proper name of the bright heavenly body you see above the west horizon just after the sunset, and 'Phosphorus' is the proper name of the bright heavenly body you see above the east horizon just before the sunrise. We initially thought that they were two distinct stars, but discovered at one point of history that Hesperus is identical with Phosphorus, which is really the planet Venus. We, of course, knew that 'Hesperus = Hesperus' is true before the discovery, but did not know that 'Hesperus = Phosphorus' is true. If the meanings of the proper names 'Hesperus' and 'Phosphorus' were exhausted by their references (extensions), then 'Hesperus' and 'Phosphorus' must mean the same thing. So if we know that Hesperus = Hesperus, then we also know that Hesperus = Phosphorus. Then, since, apparently, we knew before the astronomical discovery that 'Hesperus = Hesperus' is true, how did we not know that 'Hesperus = Phosphorus' is true before the astronomical discovery? This is called *Frege's Puzzle*.<sup>5</sup> Frege argued that we can solve this puzzle only by assuming that proper names have senses as well as references, and that 'Hesperus' and 'Phosphorus' have the same reference but different senses even though what those senses are may not be clear.

Frege, however, did not give the final word on this issue, and the intensions (or senses) of proper names have been a subject of much controversy since Frege's time. We will get back to this issue later in Section 2.8.

Going back to our main thread, the similarity between the extension and intension of a predicate, on the one hand, and the extension and intension of a singular term, on the other, is obvious. In both cases, the extension is the object(s) that satisfies (satisfy) the linguistic expression in question, whereas



the intension is how the object is (objects are) thought of. Frege famously described the sense (intension) of an expression in general as *the mode of presentation* of its reference (extension), i.e., how the reference is presented to the audience.

How about the extension and intension of a sentence? Here Frege and his followers assume the Compositionality Principle to answer this question:

- The Compositionality Principle  
The meaning of a linguistic expression is determined by the meanings of its constituents in accordance with its logical structure.

Here the meanings in question can be extension or intension. Thus, the Compositionality Principle for Extension states that the extension of a linguistic expression is determined by the extensions of its constituents in accordance with its logical structure, and the Compositionality Principle for Intension states that the intension of a linguistic expression is determined by the intensions of its constituents in accordance with its logical structure.

The ‘in accordance with its logical structure’ part will be of little relevance and may be set aside in our discussion. This qualification is needed because the extension/intension of a linguistic expression cannot be uniquely fixed simply by the extensions/intensions of the expression’s constituents regardless of its logical structure. Take ‘Adam loves Beth’ and ‘Beth loves Adam’ for instance. They obviously mean different things, so their extension or intension or both are expected to be different, but their constituents are identical: ‘Adam’, ‘Beth’, and ‘x loves y’. So the expected difference in extension/intension must be due to the difference in their logical structures, i.e., in the fact that ‘Adam’ is the subject and ‘Beth’ is the object of ‘x loves y’ in the first sentence, and the other way around in the second sentence.

Having thus eliminated the complication involving logical structures, let us now consider the sentence ‘The teacher of Alexander the Great is a philosopher’. On the one hand, since ‘the teacher of Alexander the Great’ is co-extensional with ‘the most famous disciple of Plato’, ‘the author of *Metaphysics*’, and ‘the most influential philosopher in antiquity’, all denoting the philosopher Aristotle, if we replace ‘the teacher of Alexander the Great’ with any of those co-extensional singular terms in the original sentence, the resulting sentences, ‘The most famous disciple of Plato is a philosopher’, ‘The author of *Metaphysics* is a philosopher’, and ‘The most influential philosopher in antiquity is a philosopher’, must retain the same extension as the original. On the other hand, since those singular terms have different intensions, the resulting sentences must express different intensions. Similarly, if in the sentence ‘Aristotle is a cordate’ we replace ‘cordate’ with ‘renate’, the resulting sentence ‘Aristotle is a renate’ must retain the same extension as the original while changing its intension.

So what is the extension of a sentence? What remains constant in our substitution of co-extensional expressions? It's the sentence's *truth value*, Truth or Falsity. What is the intension of a sentence? What changes in a sentence if we substitute its co-extensional but non-synonymous expressions? It's the *proposition* (or what Frege calls 'thought')<sup>6</sup> that the sentence expresses. From this line of reasoning, Frege concluded that the extension of a sentence is its truth value while the intension of a sentence is the proposition (or thought) it expresses. Recall that we have already introduced propositions in Section 1.4 as universals declarative sentences express as 0-place predicates. To clarify, propositions exist independently of the sentences that may connote (or express) them, just as properties exist independently of the predicates that may connote them. Even if we don't speak language, or even if we ourselves don't exist, properties and propositions (or Fregean 'thoughts') exist out there. We will talk a lot about propositions in the rest of this book.

Sentences in different languages may express one and the same proposition. For instance, consider:

- Snow is white;
- La neige est blanche;
- Der Schnee ist weiß;
- 雪は白い (yuki wa shiroi).

These sentences – English, French, German, and Japanese – all express one and the same proposition, that snow is white. Of course, they also have the same extension, Truth. A translation of one sentence in one language into one in another language is a correct translation if it retains the original sentence's intension as well as extension.

In summary:

Expression	Extension (denotation, or reference)	Intension (connotation, or sense)
Singular term	Individual <sup>7</sup>	Individual concept
Predicate	Set of individuals (or 'extension' in the narrow sense)	Property
Sentence	Truth value	Proposition (or Fregean 'thought')

One rather surprising consequence of this conclusion is that sentences have only two denotations: Truth and Falsity (or what Frege calls 'the True' and 'the False'). All true sentences denote one and the same thing, Truth, and all false sentences denote one and the same thing, Falsity. This is surprising at least in two respects. First, it considers Truth and Falsity genuine objects. Second, it considers completely unrelated sentences, such as 'Aristotle is a philosopher' and 'Mt Everest is the highest mountain in the world', to

denote one and the same object. In contrast, singular terms and predicates have numerous denotations. But this may not be as surprising as it is at first glance if we recall that extensions are coarse-grained and remove all differences in intension. Of course, 'Aristotle is a philosopher' and 'Mt Everest is the highest mountain in the world' connote different intensions, i.e., express different propositions; so it may not be as surprising to find out that their extensions are one and the same.

It should also be noted that, by the Compositionality Principle, if a part of a sentence does not have an extension, nor does the sentence itself. So, for instance, 'The Fountain of Youth is a hot spring' is neither true nor false (i.e., denotes neither Truth nor Falsity)<sup>8</sup> because 'the Fountain of Youth' does not have an extension (denotation).

Lastly, the Compositionality Principle is supposed to hold also between atomic and complex sentences. A typical instance of this is the truth functionality of negated, conjoined, disjoined, conditional, and biconditional sentences. As we saw in Chapter 1, the truth values of  $\neg P$ ,  $P \wedge Q$ ,  $P \vee Q$ ,  $P \rightarrow Q$ , and  $P \leftrightarrow Q$  are determined by the truth values of their constituents,  $P$  and  $Q$ . In logic we assign truth values to sentences. When we do so, we are in fact determining the sentences' extensions. Combined with what I just said in the last paragraph, if  $P$  does not have a truth value, then  $\neg P$  does not have a truth value, and if either  $P$  or  $Q$  does not have a truth value, then  $P \wedge Q$ ,  $P \vee Q$ ,  $P \rightarrow Q$ , and  $P \leftrightarrow Q$  do not have a truth value, either.<sup>9</sup> (In standard logic, however, we usually assume for the sake of simplicity that every sentence has a truth value.)

## 2.5 Possible Worlds

Many philosophers have felt for a long time that, compared to extensions, intensions are obscure and difficult to understand precisely. For instance, when are we dealing with two properties as opposed to one, and when are we dealing with two propositions as opposed to one? What are their identity conditions? How are we supposed to understand them? Are there contradictory properties such as *being a round square*? (' $x$  is a round square' has the empty extension, but does it stand for a property?) And so on ....

*Possible worlds semantics* has changed all that. Leibniz once said that this world is 'the best of all possible worlds'. So the idea of possible worlds is at least as old as Leibniz, but a possible world is philosophers' favorite toy especially in contemporary philosophy. You may have already encountered them used in various philosophical discussions; if you haven't, most likely you will. We will also talk about them at various stages of this book.

The idea of possible worlds is simple. There are numerous (probably infinitely many) possible worlds out there. Our world, the 'actual' world, is only one of those numerous possible worlds. There is nothing special about the actual world; we call it 'actual' because we live there, but the residents of

another world would call their world 'actual'. That is, 'actual' is an indexical just like 'here' and 'now'. Each possible world is slightly or radically different from the actual world or any other world. Some worlds may not have gravity, some may not even have spacetime, etc.; there also may be a possible world almost identical with the actual world except that a single atom is missing. At the same time, there are no possible worlds in which logical and mathematical laws, such as  $\neg(P \wedge \neg P)$  for any  $P$  and  $2+2=4$ , are different: there are no possible worlds in which  $P \wedge \neg P$  for some  $P$  or  $2+2=5$ . For we cannot even imagine what such possible worlds would be like. Those are *necessary* truths, i.e., truths true in all possible worlds. (We will talk about necessity and possibility, or what is called *modality*, in Chapter 3.)

More generally, each possible world is assumed to be both *maximal* (or *complete*) and *consistent*. A world is maximal (or complete) iff for any proposition  $P$ , either  $P$  holds or not- $P$  holds in that world; for instance, either there are humans or there are no humans in that world, either it rains or it does not rain in that world, etc. If a world is consistent, then for no proposition  $P$ , both  $P$  holds and not- $P$  holds; so it is not the case that both there are humans and there are no humans in that world, it is not the case that both it rains and it does not rain in that world, etc. But consistency usually means something a little more. For instance, that there are bachelors ( $P$ ) and also that everybody is married ( $Q$ ) are not *formally* inconsistent, not inconsistent in the above sense (since it is not:  $P$  and not- $P$ ); however, it is *materially* inconsistent. A world is consistent iff it is both formally and materially consistent. The actual world is assumed to be both maximal and consistent (although it is difficult to say exactly why). Since possible worlds are assumed to be similar to the actual world, it is thus usually assumed that each possible world is both maximal and consistent (even though, later in Section 3.15, we will throw some doubt on this assumption).

Some philosophers, such as David Lewis (1986), believe, literally, that there are possible worlds out there, and that our world is only one of those numerous possible worlds. This view is called *modal realism*.<sup>10</sup> Others do not believe in possible worlds, embracing *modal anti-realism*. Even many modal anti-realists, however, consider possible worlds a useful fiction and engage in possible worlds talk; this position is *modal fictionalism*. I will not try to determine in this book which position is correct, but talk as if modal realism is correct.

As I said, a possible world is philosophers' favorite toy, and you hear possible worlds talk in various areas of philosophy – metaphysics, epistemology, ethics, philosophy of mind, philosophy of science, etc. The idea, however, originates in semantics. *Semantics* is a system or the study of linguistic meaning. *Possible worlds semantics* can be used for a few different purposes, but one of them is an analysis of intension; another is an analysis of modal concepts such as necessity and possibility. I will sketch the former here and the latter in Chapter 3.

As a preliminary to our discussion on the possible worlds analysis of intension, I need to introduce and explain two concepts: *transworld identity theory* and *(mathematical) function*. The rest of this section concerns the former; the next section will discuss the latter.

There are two theories about how individuals can exist in different possible worlds: *transworld identity theory* and *counterpart theory*. Saul Kripke (1980), who contributed more than anybody else to the development of possible worlds semantics and modal logic, set forth transworld identity theory. According to transworld identity theory, literally one and the same object can exist in more than one possible world, having different, sometimes opposite, properties. For instance, according to the theory, ‘Aristotle might not have studied philosophy’ is true iff there is a non-actual possible world in which Aristotle himself, numerically one and the same man who existed in the actual world, existed but did not study philosophy.

David Lewis found this idea implausible and offered a rival theory, counterpart theory (or the theory of worldbound individuals). According to counterpart theory, no object can exist in more than one world, i.e., every object is confined to one world, or is worldbound; however, an object in a world may have a *counterpart*, a sort of twin, in another world. So, according to this theory, ‘Aristotle might not have studied philosophy’ is true iff there is a non-actual possible world in which the counterpart of Aristotle in that world, not Aristotle himself, did not study philosophy. Again, we will not decide in this book which theory of transworld identity is correct or more plausible; however, we will speak, purely for the sake of simplicity, as if transworld identity theory is correct.

## 2.6 Mathematical Functions

Another concept I need to introduce is that of *(mathematical) function*. Everybody knows at least a few mathematical functions: the addition function  $x + y$ , the multiplication function  $x \times y$ , the successor (or ‘+ 1’) function  $x + 1$ , the doubling function  $2x$ , the square function  $x^2$ , etc. A mathematical function may be thought of as a black box; you throw in certain things to the box as inputs, and it spills out certain things as outputs. The inputs are called *arguments* and the outputs are called *values* of the function. The groups of things that can be arguments and values of the function are called *the domain* and *the range (or co-domain)* of the function, respectively. So, in the case of the square function (Figure 2.12), the domain and the range are (or can be) both the set of positive integers; if the argument of the function is 1, 2, 3, 4 ..., the value of the function will be 1, 4, 9, 16 ..., respectively.

Functions can be divided into 1-place, 2-place, 3-place ... functions in a similar way in which predicates are divided into 1-place, 2-place, 3-place ... predicates. The arguments of a 1-place function are single objects; the arguments of a 2-place function are pairs of objects; the arguments of

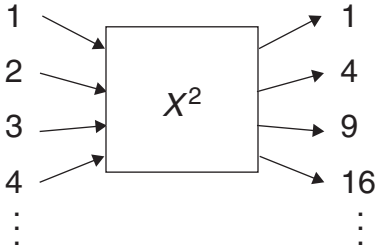


Figure 2.12 The square function.

a 3-place function are triplets of objects, etc. For instance, the successor function and the square function are 1-place functions, the addition function and the multiplication function are 2-place functions; and the function  $(x + y) \times z$  is a 3-place function. We will only be concerned with 1-place functions in the remainder of this book. So let's ignore many-place functions.

Functions can also be divided into *total* and *partial* functions. A total function gives one value in the range for every argument in the domain, whereas a partial function does not: there are arguments of a partial function which may not have corresponding values. For instance, the division function is a total function if its domain and range are rational numbers, but it is a partial function if its domain and range are positive integers:  $x/y$  always gives a rational number if  $x$  and  $y$  are rational numbers, but it does not always give a positive integer if  $x$  and  $y$  are positive integers; for instance, if  $x = 3$  and  $y = 2$ ,  $x/y$  does not give a positive integer (it gives 1.5, not an integer). Whether total or partial, a function must not give more than one value for each argument.

Finally, a *constant function* is a function whose value is one and the same (or constant) for any argument. For instance, the function that gives out the value 1 for any positive integer as an argument is a constant function.

We have so far considered only functions that have numbers in their domains and ranges. Generally, however, functions can have *any* group of things as domains and ranges. For instance, 1-place, 2-place, 3-place ... predicates are 1-place, 2-place, 3-place ... functions, respectively. The domain of a 1-place predicate, such as 'x is a dog', can be the set of individuals, and its range is the truth values, Truth and Falsity. If we supply Max or Fido as an input (argument), the predicate 'x is a dog' will give us the value Truth as the output, and if we supply Tibbles as an input, it will give us the value Falsity as the output (Figure 2.13).

The predicate 'x is or is not a dog' is a constant function that gives us Truth as the output no matter what object we supply as an input. 'The father of x' is also a function, a function, e.g., from people to people. If you put any

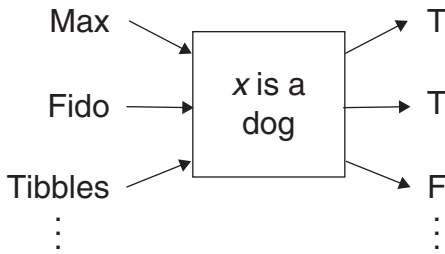


Figure 2.13 The ‘x is a dog’ function.

person as an input, it will give us his/her father as the output. Note, however, that ‘the child of  $x$ ’ is not a function because even if we put a person, say Adam, as an input, it does not give us a unique output if Adam has more or less than one child. (Alternatively, it can be regarded as a partial function whose value is missing if Adam does not have a single child.)

## 2.7 The Possible Worlds Analysis of Intension

We are now in a position to discuss the possible worlds analysis of intension. The possible worlds analysis of intension is an attempt at clarifying one previously obscure notion, that of intension, in terms of two clearer notions, those of extension and possible worlds.

The insight behind this analysis is this (Figure 2.14). We think ‘the teacher of Alexander the Great’ and ‘the most famous disciple of Plato’ connote different individual concepts (intensions) because, even though their denotations happen to be one and the same (i.e., Aristotle) in the actual world (@ below), they *could* be different; that is, we can easily imagine a possible world (W1) in which one person, say Plato, is the teacher of Alexander the Great and somebody else, say Socrates, is the most famous disciple of Plato. In that possible world, the singular term ‘the teacher of Alexander the Great’ denotes Plato and ‘the most famous disciple of Plato’ denotes Socrates.

Recall that we are assuming transworld identity theory here; so the philosophers Plato, Aristotle, and Socrates are all transworld objects depicted here as sausage-like objects penetrating different possible worlds. (Of course, there may be some other worlds in which some or all of them may not exist.)

Similarly (Figure 2.15), we think that the predicates ‘ $x$  is a cordate’ and ‘ $x$  is a renate’ connote different properties because, even though they happen to have the same extension in the actual world @, their extensions *could* be different; that is, there are possible worlds in which their extensions are different. For instance, in some possible worlds (e.g., W1 below) the set of

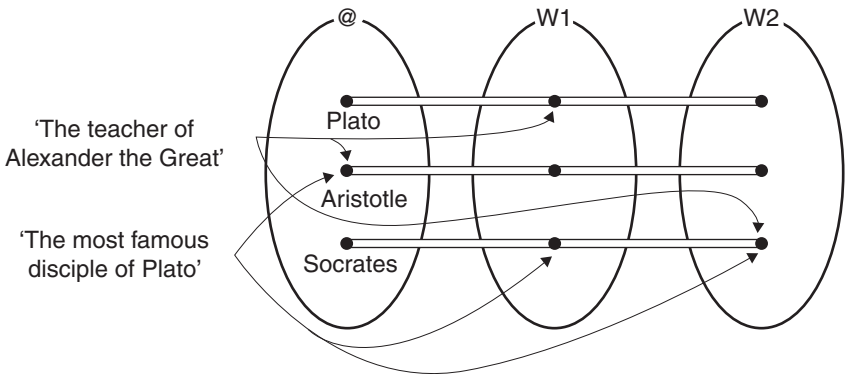


Figure 2.14 Possible denotations.

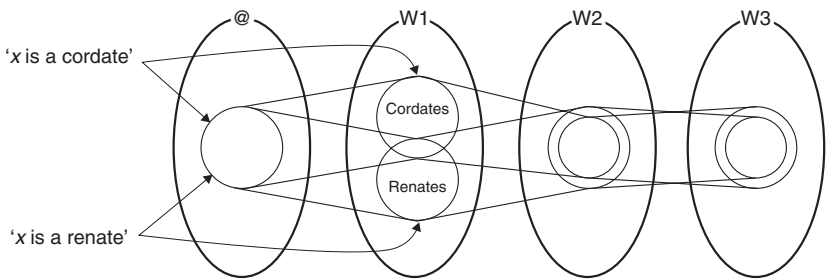


Figure 2.15 Possible extensions.

cordates and the set of renates only partially overlap; in another possible world (W2), the set of cordates is a proper subset of the set of renates; in yet another possible world (W3), the set of renates is a proper subset of the set of cordates, etc. In those worlds, 'x is a cordate' and 'x is a renate' have different extensions.

In contrast, we are inclined to think that the predicates 'x is a bachelor' and 'x is an unmarried adult male' are synonymous and express the same property. This view is nicely captured in the present way of thinking (Figure 2.16) because we cannot think of a possible world in which some bachelors are not unmarried adult males or some unmarried adult males are not bachelors; so 'x is a bachelor' and 'x is an unmarried adult male' are not only co-extensional in the actual world but co-extensional in all possible worlds (or *necessarily* co-extensional).

What is true of intensions of singular terms and predicates is also true of intensions of sentences, i.e., propositions. We think that 'Aristotle is a



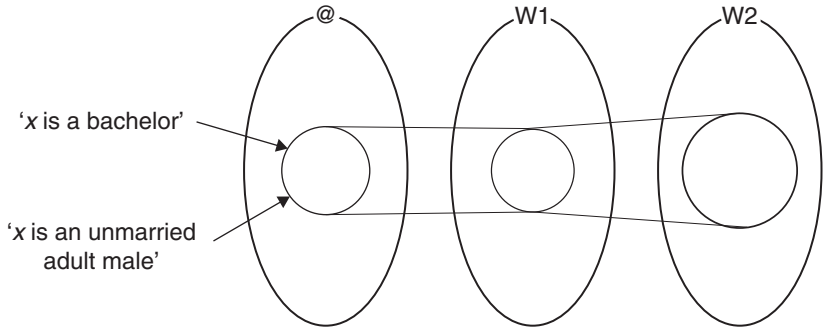


Figure 2.16 Synonymous predicates.

philosopher' and 'Mt Everest is the highest mountain in the world', though both true in the actual world, express different propositions because there are possible worlds in which Aristotle is a philosopher but Mt Everest is not the highest mountain in the world, and there are possible worlds in which Mt Everest is the highest mountain in the world but Aristotle is not a philosopher. In those worlds, 'Aristotle is a philosopher' and 'Mt Everest is the highest mountain in the world' have different truth values.

A *truth condition* of a sentence is a condition in which the sentence is true, or a necessary condition for the truth of the sentence. So a (indeed, the) truth condition of the sentence 'Aristotle is a philosopher' is that Aristotle is a philosopher, and the truth condition of the sentence 'Mt Everest is the highest mountain in the world' is that Mt Everest is the highest mountain in the world. A (not the) truth condition of 'Adam is a bachelor' is that Adam is unmarried (or male). Then the proposition a sentence expresses, understood as above, is the truth condition of the sentence, i.e., the condition in which the sentence is true.

To summarize, we can now talk about not only extensions in the actual world but extensions in other possible worlds, and the difference in intension should manifest in some world as the difference in extension in that world.

Taking up this insight, the possible worlds analysis of intension defines intension in general as follows:

- The possible worlds analysis of intension  
The *intension* of a linguistic expression is a function from each possible world to the extension of the expression in that possible world.

Here the relevant linguistic expression can be a singular term, a predicate, or a sentence. Thus:

Expression	Extension	Intension	Possible worlds analysis of intension
Singular term	Individual	Individual concept	= Function from pws to individuals
Predicate	Set (or extension)	Property	= Function from pws to extensions
Sentence	Truth value	Proposition	= Function from pws to truth values <sup>11</sup>

In particular, this analysis characterizes a proposition as follows:

- The possible worlds analysis of proposition  
The proposition a sentence connotes (or expresses) is the function from each possible world to the truth value of the sentence in that possible world.

Propositions are a major subject of our discussion in the rest of this book. When I say ‘proposition’, unless otherwise noted, I always have the above characterization in mind. As functions (which are mathematical objects), propositions exist independently of us or the sentences that may connote them.

Let me also emphasize that the intension of an expression is one constant thing (function); it does not change from world to world. What does change from world to world is the extension of the expression that the intension assigns. I’ve seen many students confused about this point.

What the above definition says, informally, is that the intension of an expression is a *rule* that determines the expression’s extension in any situation. This makes much sense from the viewpoint of language learning, too. Suppose that I am in the process of learning English, and that my teacher wants to determine if I understand the meaning of the predicate ‘*x* is a cordate’ correctly. Theoretically, if I can say correctly, for each individual in the actual world, whether it is a cordate or not, it is possible that I have understood the meaning of the predicate correctly; but there still is a chance that I cannot distinguish ‘*x* is a cordate’ from ‘*x* is a renate’. However, if the teacher somehow presents to me all possible worlds (or situations), and if I can say correctly, for each individual in each possible world, whether it is a cordate or not, I have shown that I have understood the meaning of ‘*x* is a cordate’ correctly, distinguishing it from the meaning of ‘*x* is a renate’.

The possible worlds analysis of intension analyzes intensions in terms of extensions in possible worlds. As a result, the identity conditions of intensions – when we are talking about one intension instead of two, and when we are talking about two intensions instead of one – become clearer. Contradictory predicates like ‘*x* is a round square’ still have an intension,

which is the constant function from all possible worlds to the empty extension. Overall, the possible worlds analysis has clarified the concept of intension tremendously.

## 2.8 Rigid Designators

One major issue pertaining to the concepts of extension and intension is about the intensions of proper names. A definite description (= ‘the + such and such’) has its intension ‘on its sleeves’, so to speak: it explicitly tells you what individual concept it connotes. The description ‘the teacher of Alexander the Great’ connotes the individual concept *the teacher of Alexander the Great*, and the description ‘the most famous disciple of Plato’ connotes the individual concept *the most famous disciple of Plato*. In contrast, the intension a proper name connotes is usually unclear. What individual concept does the proper name ‘Aristotle’ connote? *The teacher of Alexander the Great, the most famous disciple of Plato, the author of Metaphysics, the most influential philosopher in antiquity*, or something else? The intensions of proper names are not as clear as those of definite descriptions.

As you recall (Section 2.4), Frege maintained that proper names of one and the same object may have different intensions. For instance, ‘Hesperus’ and ‘Phosphorus’ must have the same extension, the planet Venus, but different intensions; otherwise we cannot explain why we knew before the astronomical discovery the truth of ‘Hesperus = Hesperus’ but not that of ‘Hesperus = Phosphorus’. So even though it may be unclear what intensions ‘Hesperus’ and ‘Phosphorus’ connote, they must be different, Frege concluded.

Kripke, however, argued against this view in his epoch-making book *Naming and Necessity* (1980). He introduced the concept of *rigid designator*:

- Rigid designator  
A *rigid designator* is a singular term that denotes one and the same object (individual) in all possible worlds in which the object exists.

To use the concept of function, a rigid designator is a singular term that connotes a constant (partial or total) function from all possible worlds to one and the same individual.

Then Kripke argued that proper names are rigid designators. For instance, ‘Aristotle’ is a rigid designator, denoting Aristotle in all possible worlds in which he exists. We can think of a possible world in which Aristotle was not a philosopher, did not teach Alexander the Great, did not write *Metaphysics*, or did not become the most influential philosopher in antiquity; but we cannot think of a possible world in which Aristotle is not Aristotle. So the proper name ‘Aristotle’ must denote Aristotle in other possible worlds so

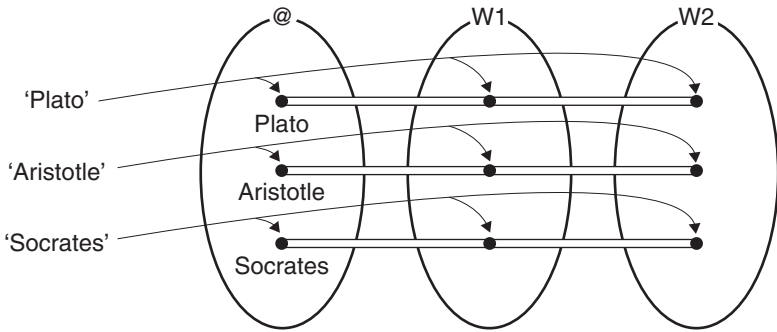


Figure 2.17 Rigid designators.

long as Aristotle exists in those worlds. Similarly, 'Plato' rigidly denotes Plato and 'Socrates' rigidly denotes Socrates (Figure 2.17).<sup>12</sup>

In contrast, definite descriptions, such as 'the teacher of Alexander the Great' and 'the most famous disciple of Plato', are typically non-rigid, as we saw in the last section.<sup>13</sup>

This is a very convincing argument, and most philosophers are in fact convinced that proper names are rigid designators. One caveat for this argument is that Aristotle may not be *called* 'Aristotle' by the residents of some other possible worlds. For instance, there may be a possible world in which Aristotle's name and Plato's name are switched, i.e., Aristotle is called 'Plato' and Plato is called 'Aristotle' by its residents. That does not make Aristotle Plato and Plato Aristotle. That world is a world in which Aristotle is still Aristotle but is called 'Plato' and Plato is still Plato but is called 'Aristotle' by its residents. Our name 'Aristotle', the name as *we* use it, still denotes Aristotle in that world even though the residents of the world call him 'Plato'.

Similarly, there are possible worlds in which the residents of the world use the word 'bachelor' differently from the way we use it in the actual world; for instance, by 'bachelor' they may just mean 'young males'. Then in *their* language, 'Every bachelor is an unmarried adult male' is not true (assuming that the other words in the sentence mean the same as they do to us). That, however, has nothing to do with the fact that in *our* language, the language we use in the actual world, 'bachelor' is synonymous to 'unmarried adult male'. Generally, the possible worlds analysis of intension is an analysis of *our* language, not the language of the residents of the other possible worlds.

Thus, again, proper names are rigid designators. It follows from this that two proper names that have the same extension (denotation) in the actual world, such as 'Hesperus' and 'Phosphorus' or 'Mark Twain' and 'Samuel Clemens', must have the same intension, too, contrary to Frege's claim (Figure 2.18).

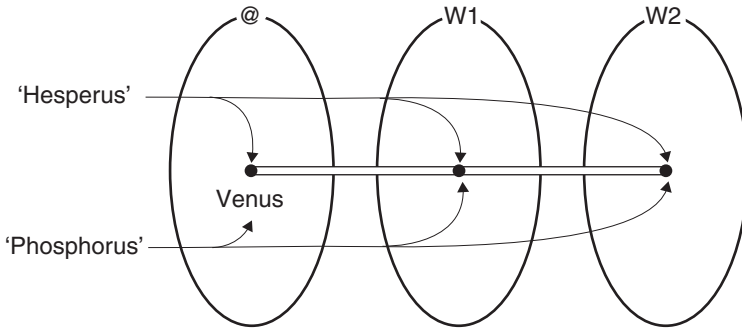


Figure 2.18 Rigid designators co-extensional in the actual world.

Note, however, that, as convincing as Kripke's argument may be, it does not give an immediate answer to Frege's original question: If 'Hesperus' and 'Phosphorus' have the same intension, how can we explain their apparent cognitive difference? People have found it difficult to answer this question from Kripke's viewpoint.

## 2.9 A Problem with the Possible Worlds Analysis of Intension

In fact, this problem about the intensions of proper names is the tip of the iceberg, an instance of a major general problem with the possible worlds analysis of intension. Put simply, the problem is that even though intensions, as characterized in the analysis, are more fine-grained than extensions (i.e., capable of drawing finer distinctions than extensions), they do not seem to be fine-grained *enough*: there are many pairs of expressions whose meanings seem to be different from each other, but the possible worlds analysis does not seem to be fine-grained enough to capture the difference.

Let me give you two more examples which should indicate the depth of the problem, one involving predicates and the other involving sentences. (As I said, we always think about singular terms, predicates, and sentences.) A *trilateral* is a figure that has three sides, just as a *triangle* is a figure that has three angles. Then the predicates '*x* is a triangle' and '*x* is a trilateral' seem to connote different properties (intensions), the first involving angles and the second involving sides. This is analogous to the cordate/renate case. However, unlike the cordate/renate case, every triangle is a trilateral and every trilateral is a triangle not only in the actual world but in all possible worlds (Figure 2.19). You cannot think of a triangle that has more or less than three sides or a trilateral that has more or less than three angles existing in any possible world, not merely in the actual world. Thus, according to the possible worlds

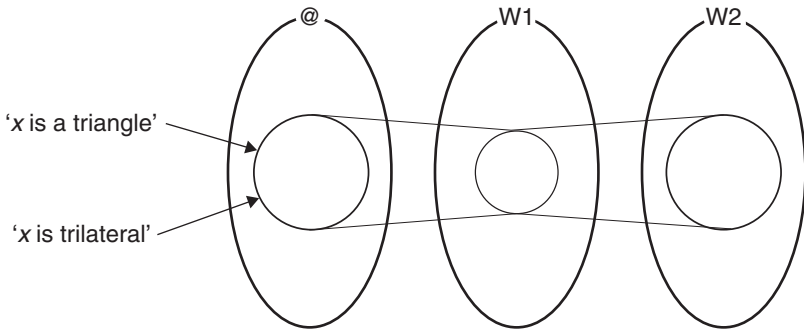


Figure 2.19 Synonymous predicates?

analysis, 'x is a triangle' and 'x is a trilateral' must be synonymous and have the same intension, contrary to what we would expect.

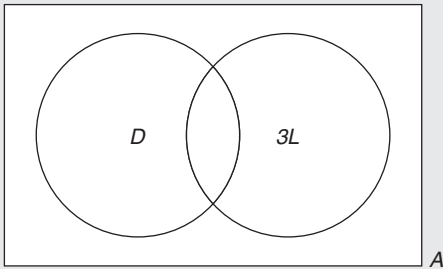
An example involving sentences may be even more shocking. As we saw, the extension of a sentence is its truth value whereas its intension is the proposition it expresses, which is identified with a function from possible worlds to truth values. So 'Aristotle is a philosopher' and 'Mt Everest is the highest mountain in the world' have different intensions even though they have the same extension, the same truth value, Truth, in the actual world. However, think of sentences that are true in all possible worlds, such as 'If it rains, then it rains', 'Every bachelor is unmarried', and ' $2+2=4$ '. Sentences that are true in all possible worlds are called *necessarily true sentences*. All the aforementioned sentences are necessarily true sentences, i.e., true in all possible worlds, for we cannot think of a possible world in which it is not the case that if it rains, then it rains, some bachelors are unmarried, or  $2+2$  does not equal 4. (Keep in mind that we are not supposed to change the meanings of the expressions involved.) So all those sentences express the same function, the constant function from all possible worlds to the truth value Truth. Hence, according to the possible worlds analysis of intension, they must express one and the same intension. But they seem to express different propositions, different thoughts. Generally, according to the possible worlds analysis, all necessarily true sentences must express one and the same intension, which is quite contrary to what we would say. Again, the analysis is not sufficiently fine-grained.

In response to many examples like these, some philosophers have proposed more fine-grained semantic concepts than intension as defined in possible worlds semantics. However, not only are those concepts much more complicated, there is no consensus as to whether they are in fact better tools

than intension. For these reasons we will not take up those new concepts and will continue using intension as a key concept in the rest of this book. I will mention some of those proposals at the end of Section 3.12.

### Exercise Questions

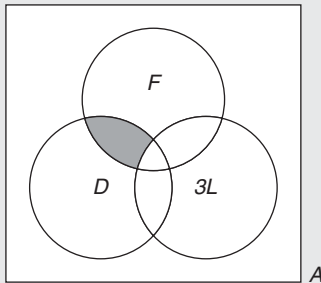
1. Explain the following concepts and distinctions.  
Sets/mereological sums (or fusions); extension (denotation, or reference)/intension (connotation, or sense); individual concepts; propositions; Compositionality Principle; possible worlds; transworld identity theory/counterpart theory; (mathematical) function; rigid designator.
2. Explain the following set-theoretic concepts.  
Empty set; unit set (or singleton); complement; intersection (or product); union (or sum); subset; proper subset.
- \*3. Which part of the Venn diagram does each of the following expressions denote? Shade the area. Keys. *D*: dogs; *3L*: three-legged; *A*: animals.



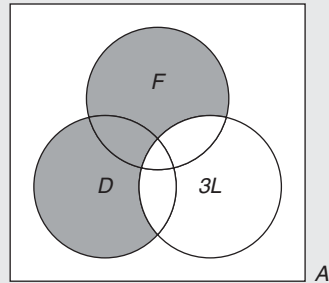
- (a) The set of dogs.
- (b) The set of animals except dogs.
- (c) The set of three-legged dogs.
- (d) The set of three-legged animals that are not dogs.
- (e) The set of dogs and three-legged animals.
- (f) The set of animals that are neither dogs nor three-legged.

Describe the set denoted by the shaded area(s) of each of the following Venn diagrams. An additional key. *F*: female. (You may assume that all animals are either male or female.)

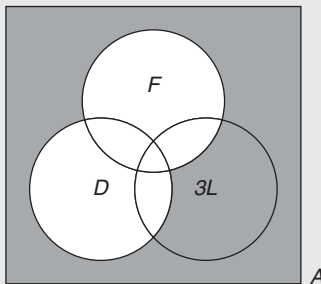
(g)



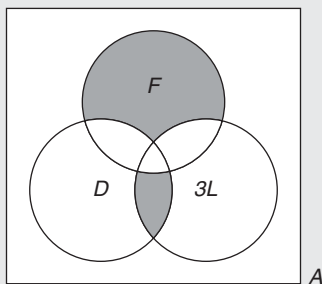
(h)



(i)



(j)



4. How is the set-membership relation in set theory different from the part-whole relation in mereology?
5. What are the extension and intension of a linguistic expression in general, and the extension and intension of a singular term, a 1-place predicate, and a sentence, in particular? How are those extensions (intensions) related to one another? (Mention the Compositionality Principle.)
- \*6. Name or describe a function that would return the values on the right of the arrow from the arguments on the left.
  - (a)  $4 \rightarrow 3; 3 \rightarrow 2$ .
  - (b)  $2 \rightarrow 8; 4 \rightarrow 64$ .
  - (c)  $9 \rightarrow -18; -18 \rightarrow 36$ .
  - (d)  $\text{Socrates} \rightarrow \text{Plato}; \text{Plato} \rightarrow \text{Aristotle}$ .
  - (e)  $\text{Donald Trump} \rightarrow \text{Barack Obama}; \text{Barack Obama} \rightarrow \text{George W. Bush}$ .
  - (f)  $\text{Aristotle} \rightarrow \text{T}; \text{Donald Trump} \rightarrow \text{F}$ .
7. Describe the possible worlds analysis of intension. What is the main problem with the analysis?



### Discussion Question

1. Claim: for any set  $S$ , a proper subset of  $S$  has fewer members than  $S$ . Discuss whether this claim is true or false. Hint: Compare the set  $N$  of natural numbers,  $0, 1, 2, 3 \dots$  with the set  $E$  of even natural numbers,  $0, 2, 4, 6 \dots$ .  $E$  is a proper subset of  $N$ . However, there is one-one correspondence between the members of  $N$  and the members of  $E$ ; just double the numbers in  $N$ , and you will obtain the numbers in  $E$ . So ....

### Suggested Further Reading

You don't have to know technical details of set theory or mereology to do philosophy (except philosophy of mathematics). If you are interested, you can read and benefit from, for instance:

- Patrick Suppes, *Axiomatic Set Theory*.
- Roberto Casati and Achille Varzi, *Parts and Places*.

The two most important writings this book centers around are:

- Gottlob Frege, "Über Sinn und Bedeutung (On Sense and Reference)" (several English translations available).
- Saul Kripke, *Naming and Necessity*.

Fortunately, these are also very accessible even to undergraduate students. You can benefit tremendously from reading at least the first one-third of the former and the entirety of the latter.

- Joseph LaPorte, "Rigid Designators"

is a thorough survey of the theories of rigid designators.

### Notes

- 1 This word, as well as 'renate', is a very technical term most smaller dictionaries do not contain. They may instead contain the word 'chordate', but chordates are animals with notochords and are totally unrelated to cordates, discussed here.
- 2 Descartes called primary qualities such as sizes and shapes of objects 'extensions' because those qualities are *extended* in spacetime. It is best to think of these two senses of 'extension' as unrelated.
- 3 However, mereology is not restricted to spatiotemporal objects. A mereological theory can be constructed in any field in which 'part-whole' talk makes sense,

including fields that deal with abstract objects. For instance, the common noun ‘professor’ is part of the predicate ‘is a professor’, which is part of the sentence ‘Adam is a professor’; so ‘professor’ is also part of the sentence. Here we are talking about not expression tokens but expression types, which are abstract objects. But a mereological theory can be constructed for those grammatical parts.

- 4 The polar opposite of mereological universalism, the view that two mereological atoms never compose another object, is called *mereological nihilism*. As you can easily imagine, one question that immediately arises for mereological nihilism is: What is a mereological atom?
- 5 By Frege’s Puzzle, some (probably many) philosophers refer to the puzzle about the cognitive differences between ‘ $a = a$ ’ and ‘ $a = b$ ’, where  $a$  and  $b$  are singular terms in general, i.e., not only proper names but including definite descriptions such as ‘the teacher of Alexander the Great’ and ‘the most famous disciple of Plato’. Others include in Frege’s Puzzle even the puzzle about propositional attitudes we will discuss in Section 4.3.
- 6 So Fregean thoughts exist even when nobody is holding those thoughts.
- 7 I feel obligated to add a note here, which may be ignored if you are still in the process of digesting the basics. Consider, for instance, the singular terms “the intension of the expression ‘the teacher of Alexander the Great’ (or ‘ $x$  is a philosopher’ or ‘The teacher of Alexander the Great was a philosopher’).” The extensions of these singular terms are intensions (of the singular term ‘the teacher of Alexander the Great’, of the predicate ‘ $x$  is a philosopher’, and of the sentence ‘The teacher of Alexander the Great was a philosopher’, respectively). Generally, an intension of an expression can be treated as an individual, the extension of some singular term (which has its own intension). In the list, the term ‘individual’ is used more broadly than before, including such intensions taken as individual objects. Any kind of object can be taken as an individual in this sense. I’ve noted a similar problem (Chapter 1, note 12) in which abstract nouns, instead of predicates, seem to denote universals.
- 8 Bertrand Russell and his followers may say that this sentence is false. However, I cannot get into a discussion about why, because their view is (usually) based on Russell’s theory of definite descriptions, but I have decided not to discuss the theory and its implications (Chapter 1, note 8).
- 9 Some logicians found this counterintuitive and developed a logic (or logics) according to which even if  $P$  is neither true nor false,  $P \rightarrow P$  and  $P \vee \neg P$  are still true,  $P \wedge Q$  is false so long as  $Q$  is false, and  $P \vee Q$  is true so long as  $Q$  is true (Kleene 1952; van Fraassen 1968). In this book, however, we will not employ such a logic and take a complex sentence as neither true nor false if one of its constituent atomic sentences is neither true nor false.

The negation is a tricky device. They can be ambiguous. Consider ‘The Fountain of Youth is *not* a hot spring’ or ‘The present King of France is *not* bald’ (Russell’s famous example). ‘The Fountain of Youth is not a hot spring’ can be understood in two ways. It can be understood as ‘The Fountain of Youth is not-a-hot-spring (i.e., cold water)’, or it can be understood as ‘It is not true that the Fountain of Youth is a hot spring’. The former is neither true nor false because

the Fountain of Youth does not exist. The latter is true because ‘The Fountain of Youth is a hot spring’ is certainly not true (though it is not false, either). Similarly, ‘The present King of France is not bald’ can be understood as ‘The present King of France is non-bald (i.e., hirsute)’ or ‘It is not true that the present King of France is bald’. The former is neither true nor false because France is a republic and no longer has a King; the latter is true because ‘The present King of France is bald’ is not true (or false). The negation used in the former way, denying only the relevant predicates, is called *the internal negation*; the negation used in the latter way, denying the whole sentences, is called *the external negation*.

- 10 A watered-down version of modal realism is called *ersatz realism* (or *ersatzism*) (Carnap 1947, Plantinga 1972, Adams 1974, Stalnaker 1984). ‘Ersatz’ means ‘substitute’ or ‘fake’ in German. Ersatzism accepts the existence of possible worlds on surface but takes a possible world as something very different from what the name suggests, as something you can construct from things existing in the actual world; for instance, one version of ersatzism takes a possible world as a maximal state of affairs existing in the actual world; another takes it as a maximal consistent set of propositions; yet another takes it as a maximal consistent sentence, etc. It may seem a little misleading to categorize ersatzism as a version of modal realism, but that’s how philosophers do it.
- 11 Many philosophers instead say that the intension of a sentence is the set of possible worlds in which the sentence is true. This alternative characterization is acceptable in the context in which we may assume that every sentence is either true or false and not valueless in every possible world; for then any set of possible worlds in which the relevant sentence is true uniquely determines a function from possible worlds to truth values, and vice versa. However, in the context in which this assumption cannot be made, that characterization is inaccurate because a set of possible worlds in which the sentence is true does not uniquely determine a function from possible worlds to truth values: there are functions that give truth value Truth to the sentence in those, same, possible worlds, but are different in assigning Falsity to the sentence in different possible worlds. So the characterization given in the main text here is better.
- 12 John Stuart Mill held a view similar to Kripke’s:

Proper names are not connotative; they denote the individuals who are called by them, but they do not indicate or imply any attributes as belonging to those individuals.

(Mill 1843, Book I, Chapter 2, Section 5)

Because of this historical precedent, proper names that do not work like definite descriptions are often called *Millian names*, and the view that all proper names are Millian names is often called *the Millian view*.

- 13 This does not mean there aren’t any rigid definite descriptions; there are. We will see some examples in Sections 3.7 and 5.6.

# Analyticity, Apriority, and Necessity

---

### 3.1 Four Distinctions in Truths

We will begin this chapter by making four distinctions in truths (i.e., true sentences or propositions): logical vs non-logical truths, analytic vs synthetic truths, a priori vs a posteriori truths, and necessary vs contingent truths. Necessity, possibility, and contingency are called *modalities*. We will investigate the relations between those four distinctions, but gradually move toward a discussion about the modal concepts and the possible worlds analysis of those concepts.

In what follows we will talk about the truth and falsity of propositions as well as sentences. Recall that a proposition, generally, is a function from possible worlds to truth values. We define *a proposition true/false in possible world* as follows:

- *A proposition is true in possible world  $w$*  iff it is a function from possible worlds to truth values that gives out Truth as its value if its argument is  $w$ .
- *A proposition is false in possible world  $w$*  iff it is a function from possible worlds to truth values that gives out Falsity as its value if its argument is  $w$ .

Then the relation between a true sentence and a true proposition will be as follows:

- A sentence is true in possible world  $w$  iff
  - (a) it expresses a proposition; and
  - (b) that proposition is true in  $w$ .

Analogously for falsity. Since, as I've said several times, propositions exist independently of us or the sentences that may express them while sentences are human creations, the above biconditional may be taken as a two-step way for a sentence to be true: a sentence is true by, first, expressing a proposition, and, second, the proposition's being true. In this picture, propositional truth

is primary and original, and sentence truth is secondary and derivative. We will often drop the phrase ‘in the actual world’ from ‘a proposition true/false in the actual world’ and simply say ‘a true/false proposition’ (or sometimes ‘an *actually* true/false proposition’).

The next few sections will present the four distinctions as well as pertinent key definitions such as follows:

- (1) Logical vs non-logical truths – a logical distinction.
  - A *logical truth* is a sentence true purely by virtue of the meanings of the logical connectives involved, independently of the facts in the world.
- (2) Analytic vs synthetic truths – a semantic distinction.
  - An *analytic truth* is a sentence true purely by virtue of the meanings of the words involved, independently of the facts in the world.
- (3) A priori vs a posteriori truths – an epistemic (or epistemological) distinction.
  - An *a priori truth* is a true proposition that can be known a priori.
- (4) Necessary vs contingent truths – a metaphysical distinction.
  - A *necessary truth* is a true proposition that could not be false.

If you have taken at least one logic course, you should be more or less familiar with the concept of logical truth. As I said in Chapter 2, semantics is a system or the study of meaning. So the semantic distinction between analytic and synthetic truths is a distinction involving the meanings of words. *Epistemology* is the study and theory of knowledge. So the a priori/a posteriori distinction has something to do with how to obtain the relevant knowledge. *Metaphysics* is the study and theory of the nature of things. Modalities include necessity, possibility, and contingency, and how things can or cannot be, i.e., the essence of things, is a metaphysical issue. So the modal distinction between necessary and contingent truths is more broadly a metaphysical distinction. Possible worlds will turn out to be a useful device for the analysis of modal concepts, as it was for the analysis of intension.

Now we will get to a detailed examination of the four distinctions.

## 3.2 Logical vs Non-logical Truths

A logical truth is a sentence true purely by virtue of the meanings of the logical connectives involved, independently of the facts in the world. The relevant logical connectives here include negation  $\neg$ , conjunction  $\wedge$ , disjunction  $\vee$ , conditional  $\rightarrow$ , and biconditional  $\leftrightarrow$  in propositional logic and the universal and the existential quantifier,  $\forall x$  and  $\exists x$ , in predicate logic. Logical truths in propositional logic in particular are also called *tautologies*.

Again, suppose  $P$ : Adam is a professor and  $Q$ : This is a university building. Then the following are all logical truths:

- $P \vee \neg P$ : Adam is a professor or not a professor.
- $\neg(P \wedge \neg P)$ : it is not the case that Adam is both a professor and not a professor.
- $P \rightarrow P$ : if Adam is a professor, then Adam is a professor.
- $(P \wedge Q) \rightarrow P$ : if Adam is a professor and this is a university building, then Adam is a professor.
- $P \rightarrow (P \vee Q)$ : if Adam is a professor, then either Adam is a professor or this is a university building.

These sentences are true regardless of the truth values of  $P$  and  $Q$  purely by virtue of the meanings of the relevant connectives, represented by the truth tables for the connectives. For instance, applying the truth tables for negation, disjunction, and conditional presented in Section 1.3, the truth tables for  $P \vee \neg P$  and  $P \rightarrow P$  will look like this:

$P$	$P$	$\vee$	$\neg$	$P$
T	T	<b>T</b>	F	T
F	F	<b>T</b>	T	F

$P$	$P$	$\rightarrow$	$P$
T	T	<b>T</b>	T
F	F	<b>T</b>	F

(The first and the second row of the first table say: if  $P$  is T, then  $\neg P$  is F, so  $P \vee \neg P$  is T; if  $P$  is F, then  $\neg P$  is T, so  $P \vee \neg P$  is T. Similarly for the second table.) As you can see here,  $P \vee \neg P$  and  $P \rightarrow P$  are true regardless of whether  $P$  is true or false. So  $P \vee \neg P$  and  $P \rightarrow P$  are both logical truths. The principle that for any sentence  $P$ ,  $P \vee \neg P$  is a logical truth is called the Law of Excluded Middle, as it asserts that there is nothing between  $P$  and not- $P$ . An alternative definition of logical truth is that a logical truth is a sentence true purely by virtue of its logical structure.

A logical truth, i.e., a logically true sentence, expresses a proposition true in all possible worlds, which, of course, is true also in the actual world. Then, to take the aforementioned two steps a sentence must take in order to be true, a logical truth is true in the actual world.

As I said in Section 1.4, the identity predicate ' $x = y$ ' is often considered a logical operator along with the connectives and quantifiers. If it is, then sentences such as ' $a = a$ ', ' $a = b \rightarrow b = a$ ', and ' $(a = b \wedge b = c) \rightarrow a = c$ ' for any singular terms  $a$ ,  $b$ , and  $c$ , may also be considered logical truths.

Suppose that  $P$  (Adam is a professor) is actually true. Then  $P \vee Q$  (Adam is a professor or this is a university building) and  $P \vee \neg Q$  (Adam is a professor or this is not a university building) are also true. But neither of these is logically true. They are non-logical truths. So are all true atomic sentences, i.e., sentences not containing any of the logical connectives, such as 'Earth is round' and 'Aristotle is a philosopher'.

The mirror image of a logical truth is a logical falsity. A *logical falsity* is a sentence false purely by virtue of the meanings of logical connectives involved. Examples are  $P \wedge \neg P$ ,  $\neg(P \rightarrow P)$ , and  $P \leftrightarrow \neg P$  for any  $P$ . Clearly, the negation of a logical truth is a logical falsity, and the negation of a logical falsity is a logical truth. A logical falsity may also be called a *contradiction*. The principle that for any  $P$ ,  $\neg(P \wedge \neg P)$  is a logical truth is called the Law of Non-Contradiction, as it asserts that there is no contradiction in the world.

### 3.3 Analytic vs Synthetic Truths

An analytic truth is a sentence true purely by virtue of the meanings of the words involved, independently of the facts in the world. The opposite of analytic truths is synthetic truths; so a synthetic truth is a sentence true at least partly by virtue of the facts in the world.

Logical truths are a species of analytic truths: every logical truth is analytically true. This is clear from the definitions of analytic truth and logical truth; an analytic truth is a sentence true by virtue of the meanings of the words involved, but if those words are in particular logical connectives, that analytic truth will also be a logical truth. There are, however, analytic truths that are not logically true, such as conceptual truths and mathematical truths, listed below with examples:

- Analytic truths
  - Logical truths. See above.
  - Conceptual truths. E.g., Every bachelor is unmarried. A vixen is a female fox. An ox is a male cow. An actress is a female actor. Every triangle has three angles.
  - Mathematical truths. E.g., Every triangle has three sides. There is exactly one empty set.  $2+2=4$ .

For instance, 'bachelor' simply means 'unmarried adult male'. Thus, the truth of the sentence 'Every bachelor is unmarried' follows simply from the concepts of 'bachelor' and 'unmarried', regardless of what the world is like. (This sentence is true even if there are no bachelors in the world, for it only says that *if*  $x$  is a bachelor,  $x$  should be unmarried.) Similarly, 'triangle' simply means 'a figure with three angles'. The truth of the sentence 'Every triangle has three angles' simply follows from this. The distinction between

conceptual truths and mathematical truths is blurry; but the basic idea here is that mathematical truths follow from the definitions of the concepts involved.<sup>1</sup> Every triangle, by definition, is a three-angled figure. But a simple reflection tells us that every three-angled figure is also a three-sided figure. It follows that every triangle has three sides. An empty set, by definition, is a set with no member. But, by the Axiom of Extensionality, there cannot be two distinct sets with no members. Therefore, there is exactly one empty set. Similarly, ' $2+2=4$ ' follows from the concepts of ' $2$ ', ' $4$ ', addition, and identity. You prove mathematical truths; those proofs are ways to show that these truths follow from those original definitions (called 'axioms'). The name 'analytic truths' comes from the fact that we can obtain those truths simply by analyzing the relevant concepts, meanings, or definitions.

The opposite of analytic truths is synthetic truths. To obtain synthetic truths, analyzing concepts is not enough; we must synthesize (or incorporate) actual facts in the world. Suppose every bachelor is in fact secretly hoping for marriage. Even if that's the case, the truth of the sentence 'Every bachelor hopes for marriage' does not simply follow from the concept of 'bachelor' and 'marriage'; we must launch a scientific investigation to obtain that truth. 'Triangles are Euclid's favorite figures' – even if that is so, it does not follow from the relevant concepts alone.

Finally, before moving on to the other two distinctions in truths, the a priori/a posteriori distinction and the necessary/contingent distinction, I would like to introduce a couple of derivative definitions. The concepts of logical truth and analytic truth apply primarily to sentences, not propositions. This should be clear from the definitions of logical truth and analytic truth, both of which mention 'meanings of words': 'An analytic truth (a logical truth) is true purely by virtue of the meanings of the (logical) words involved'. Propositions don't have meanings; they *are* the meanings of sentences. However, as you will see immediately below, since the concepts of apriority and necessity apply primarily to propositions, it will be convenient to introduce the concepts of logically true *proposition* and analytic *proposition*. Many philosophers mean 'logically true *propositions*' and 'analytic *propositions*' when they talk about logical truths and analytic truths. I define a logically true proposition and an analytic proposition as follows:

- A *logically true proposition* is a proposition expressed by some logically true sentence.
- An *analytically true proposition* is a proposition expressed by some analytically true sentence.

A not-logically-true proposition can be defined as the opposite of a logically true proposition, and a synthetically true proposition as the opposite of an analytically true proposition. Logically true propositions and analytically true propositions may also be called 'logical truths' and 'analytic truths' so



long as there is no confusion between these and logically true and analytic sentences, respectively.

### 3.4 A Priori vs A Posteriori Truths

The distinction between the a priori and the a posteriori is originally a distinction in the kinds of knowledge. In the modern era, Continental Rationalists, such as Descartes, Spinoza, and Leibniz, argued that we can obtain knowledge from reasoning alone without having any experience. In contrast, British Empiricists, such as Locke, Berkeley, and Hume, maintained that all knowledge must originate in sense experience (although they did not deny that logical and mathematical knowledge can be obtained from the reasoning about the relations between ideas alone once those ideas are obtained empirically). ‘A priori’ means in Latin ‘prior to’ (i.e., ‘before’) and ‘a posteriori’ means ‘posterior to’ (i.e., ‘after’). Here ‘a’ is not an indefinite article (as in ‘a tree’); it’s best to treat ‘a priori’ and ‘a posteriori’ as single words. But before and after what? Experience. So a priori knowledge is knowledge we can obtain before (or without) any experience, and a posteriori knowledge is knowledge we can obtain only after (or with) some experience.

An a priori truth, then, is a true proposition we can obtain as a priori knowledge, and an a posteriori truth is a true proposition we can obtain only as a posteriori knowledge.

We will talk about knowledge thematically later in Chapter 5, Part B; but, for now, I would like to say three things about knowledge in general and a priori knowledge in particular. First, the object of knowledge is propositions, not sentences. Consequently, an a priori truth ought to be a proposition, not a sentence. The object of knowledge ought not to be sentences because, among other things, animals and infants, who do not speak language, can still obtain knowledge. A dog or a baby seems to know that the dinner is ready or that their favorite toy is in the box. We will discuss the objects and content of our mental states such as beliefs, desires, and knowledge in Chapter 4; but it is at least initially plausible to think that the object of our knowledge is not sentences but propositions which, recall, exist independently of the sentences which may express them.

Second, it’s not as if we don’t need any experience *whatsoever* to obtain a priori knowledge. Reasoning is an experience in a broad sense; so is the understanding of the concepts, meanings, and definitions involved. A priori knowledge is knowledge we can obtain as soon as we have those experiences, without any further fact-checking by, e.g., perception or memory.

Lastly, an a priori truth *can* be obtained a priori, but it *need not* be; it can be obtained a posteriori. For instance, you can obtain the mathematical knowledge that there is exactly one empty set a priori, by reasoning from the Axiom of Extensionality; but you can obtain the same knowledge a

posteriori, too, for instance by hearing from your math professor and simply taking his word for it.

What are instances of a priori truths? How can we know anything a priori? Here it seems reasonable to think that the proposition expressed by any analytic sentence, such as ‘Adam is a professor or not a professor’, ‘Hesperus is identical with Hesperus’, ‘Every bachelor is unmarried’, ‘Every triangle has three sides’, ‘There is exactly one empty set’, or ‘ $2+2=4$ ’, can be known a priori. For we can know it simply by understanding the meanings of the words involved without any substantive experience. We will consider the relations between the different kinds of truth in Section 3.10, but at this point it seems reasonable to generalize that all analytically true sentences express a priori truths.

A posteriori truths are not difficult to find. Most truths about the world we obtain directly or indirectly from sense experience are a posteriori. For instance, propositions such as that snow is white, that the sky is blue, that I am now looking at the computer screen, that Earth is round, that Mt Everest is the highest mountain in the world, and that light travels at 300,000 km/s are a posteriori truths.

### 3.5 The Possible Worlds Analysis of Modality; Modal Logic

The last of the four distinctions in truths is the distinction between necessary truths and contingent truths; but this distinction can be better understood as a combination of three-way *modal* distinction between necessary, contingent, and impossible propositions, on the one hand, and the distinction between actually true and actually false propositions, on the other. A contingent proposition can be actually true or actually false. (For the sake of simplicity, in this section I will assume that every proposition is either true or false in each possible world.) A contingent truth is a contingent proposition that is actually true. An actual truth is either a necessary truth or a contingent truth.

Possible		Impossible
Necessary	Contingent	
Actually true		Actually false

It may not be surprising by now, but modal concepts such as necessity, possibility, and contingency can be analyzed in terms of possible worlds. This analysis may be called *the possible worlds analysis of modality*. The following are the definitions and possible worlds analyses of key modal concepts, where *P* is any proposition:

- *P* is possible iff *P* can be true iff there is at least one possible world in which *P* is true.

- $P$  is impossible iff  $P$  cannot be true iff there is no possible world in which  $P$  is true.
- $P$  is necessary iff  $P$  must be true iff  $P$  is true in every possible world.
- $P$  is contingent iff  $P$  can be true and can be false iff there is at least one possible world in which  $P$  is true, and there is at least one possible world in which  $P$  is false.
- $P$  is contingently true iff  $P$  is contingent and is actually true.
- $P$  is contingently false iff  $P$  is contingent and is actually false.

A contingent proposition is so called because its truth value in each possible world is contingent on the facts in the world.

The logic that deals with modality is called *modal logic*. In modal logic, ‘ $P$  is necessary’ and ‘ $P$  is possible’ are symbolized as  $\Box P$  and  $\Diamond P$ , respectively, where  $\Box$  and  $\Diamond$  are called ‘box’ and ‘diamond’.<sup>2</sup> Again, in this section, we will assume that  $P$  is either true or false in each possible world. Then the following biconditionals hold for any  $P$ :

- $\neg\Box P$  (It is not the case that  $P$  is necessary) iff  $\Diamond\neg P$  (not- $P$  is possible);
- $\neg\Diamond P$  (It is not the case that  $P$  is possible, i.e.,  $P$  is impossible) iff  $\Box\neg P$  (not- $P$  is necessary).

These are so because

- $\Box P$  iff  $\forall w(P \text{ is true in } w)$ , i.e.,  $P$  is necessary iff for every possible world  $w$ ,  $P$  is true in  $w$ ;
- $\Diamond P$  iff  $\exists w(P \text{ is true in } w)$ , i.e.,  $P$  is possible iff for some possible world  $w$ ,  $P$  is true in  $w$ .

So  $\neg\Box P$  iff  $\neg\forall w(P \text{ is true in } w)$  iff  $\exists w\neg(P \text{ is true in } w)$  by a Generalized De Morgan’s Law (Section 1.4), and this is so iff  $\Diamond\neg P$ . Similarly for the second equation. (You may call the above equations ‘Modal De Morgan’s Laws’.) Generally, necessity  $\Box$  can be taken as truth in all possible worlds ( $\forall w$ ), and possibility  $\Diamond$  can be taken as truth in at least some possible world ( $\exists w$ ). The relation between box  $\Box$  and diamond  $\Diamond$  is analogous to the relation between the universal quantifier  $\forall x$  and the existential quantifier  $\exists x$ . Just as in the quantifier case,  $\Box$  and  $\Diamond$  are mutually definable:  $\Box P = \neg\Diamond\neg P$  and  $\Diamond P = \neg\Box\neg P$ . So if we have one, we can introduce the other by definition.

Another important set of relations is the relations between necessary, possible, and actual truths:

- If  $\Box P$  ( $P$  is necessary), then  $P$  is actually true;
- If  $P$  is actually true, then  $\Diamond P$  ( $P$  is possible); thus,
- If  $\Box P$  ( $P$  is necessary), then  $\Diamond P$  ( $P$  is possible).

These relations can be read off easily from the table above.

### 3.6 Metaphysical Modality; the Necessary Truth

Necessity and possibility are often expressed by the auxiliary verbs ‘must’ and ‘can’, respectively. But just as there are different uses of ‘must’ and ‘can’, there are different kinds of modality. For instance, consider:

- I don’t see Adam. He must be coming late.
- You must help people in need.
- A triangle must have three sides.

The first ‘must’ expresses epistemic certainty and the second expresses moral duty; they are called *epistemic modality* and *deontic modality*, respectively. In comparison, the third ‘must’ expresses the nature and essence of things, in this case triangles. Such modality is called *metaphysical modality*. We will discuss the first two kinds of modality briefly later in Section 3.15; but in this chapter we are mainly interested in metaphysical modality. The working assumption here is that things have objective modal characteristics on their own, quite independently of what we think of them.

What are examples of metaphysically necessary truths? Here again, analytic truths – logical, conceptual, and mathematical truths – seem to be the prime candidates for necessary truths. Those truths are not contingent on the physical conditions of the world. Even if there are no material objects in the world, if it is raining, then it is raining,  $2+2=4$ , and all bachelors are unmarried.

There is a slight complication here, however. Take for instance ‘Adam is either a professor or not a professor’. Even though there is no possible world in which this sentence is false, there are possible worlds in which the sentence is not true, either, because Adam does not exist in those worlds. (We ignored this possibility in the last section when we assumed that every proposition is either true or false.) To cancel out the possibility of this kind, we often add the qualification ‘if the relevant objects exist’; so instead of saying ‘Adam is a professor or not a professor’, we say ‘If Adam exists, Adam is a professor or not a professor’. Then this conditional sentence will be true in all possible worlds; it will express a necessary truth. Similarly, the sentence ‘Hesperus is identical with Hesperus’ does not express a necessary truth; but ‘If Hesperus exists, Hesperus is identical with Hesperus’ does.

One important fact about necessary truths is that, in fact:

- There is exactly one necessary truth (or necessarily true proposition).

This is so because there is only one constant function that gives out the truth value Truth for any possible world as the argument. So we can talk about ‘the necessary truth’ instead of ‘a necessary truth’ (as I just did) or ‘necessary truths’.

In addition to analytic truths, some sentences expressing the essences of things are also usually considered to express the necessary truth. For example, consider 'If water<sup>3</sup> exists, water is  $H_2O$ ' (i.e., 'If water exists, water =  $H_2O$ ') and 'If heat exists, heat is mean molecular kinetic energy' (i.e., 'If heat exists, heat = mean molecular kinetic energy'). Note that when we say 'Water =  $H_2O$ ', we are treating 'water' as a proper name (of a substance) and ' $H_2O$ ' as an abbreviation of a definite description, something like 'dihydrogen monoxide' or 'the mereological sum of molecules each of which consists of two hydrogen atoms and one oxygen atom'; similarly for 'heat' and 'mean molecular kinetic energy'. 'If water exists, water is  $H_2O$ ' and 'If heat exists, heat is mean molecular kinetic energy' express the necessary truth because in any possible world in which there is water/heat, it is  $H_2O$ /mean molecular kinetic energy. We will expand on essence in the next section.

By the way, notice that the definite descriptions ' $H_2O$  (= the mereological sum of molecules each of which consists of two hydrogen atoms and one oxygen atom)' and 'mean molecular kinetic energy' (though the latter does not have the usual 'the') are rigid designators, for they denote water/heat in all possible worlds in which it exists. Thus, even though most definite descriptions are non-rigid designators, the definite descriptions describing essences are rigid.

### 3.7 Essence and Haecceity

Essence has been a subject of philosophical debate since Aristotle first introduced the concept. An *essential property* of an individual or a group (or a kind) of individuals is a property that makes it what it is. It must have that property to exist; if it loses it, it no longer will be the same thing, and becomes something else. So an essential property for  $x$  is a property necessary for the existence of  $x$ . In the possible worlds terminology, there is no possible world in which  $x$  exists without having that property. A property which is not an essential property is an *accidental property*. The *essence* of an individual or a group of individuals  $x$  is a property or a set of properties that is both necessary and sufficient for the existence of  $x$ . *Haecceity* (pronounced hek'si:iti, originally Latin, meaning 'thisness') is the essence of an individual, as opposed to a group of individuals.

Thus, for instance, the property *being an animal* is an essential property of a dog, for a dog cannot exist without being an animal. In contrast, the property *having four legs* is an accidental property of a dog, for a dog can exist without having four legs. *Being  $H_2O$*  is an essential property – and probably the essence – of water.

The haecceity of an individual person is more controversial. What makes Aristotle Aristotle and nothing else? His distinctive DNA? Kripke said that we could not have been born from anybody but our actual parents; if our parents had been different, we would not have been us. If that's true, then

our origin must be our essential property. It cannot be our whole essence, though, for our biological siblings have the same property but still are not us. *Being a philosopher* and *being the teacher of Alexander the Great* are not Aristotle's essential properties but his accidental properties, for Aristotle could have never studied philosophy or could have never been in Alexander's court and could still have been Aristotle.

The distinction between essential and accidental properties has an analogue in relations: the distinction between *internal* and *external relations*. Just as  $x$  cannot exist without having its essential properties,  $x$  and  $y$  cannot exist without having internal (2-place) relations between them. (Similarly for 3-place relations between  $x$ ,  $y$ , and  $z$ , etc.) Internal relations are usually considered to arise from essential properties of the objects involved ( $x$ ,  $y$ , etc.). For instance, suppose (in an unlikely scenario) that *being 6 feet tall* is Adam's essential property while *being 5 feet tall* is Betty's essential property. Then *Adam's being taller than Betty* is an internal relation between Adam and Betty: Adam and Betty cannot exist without Adam's being taller than Betty. External relations are relations that are not internal. Some lovers think that their being in love with each other is an internal relation; given the essential properties about their characters, etc., it was impossible for them not to fall in love with each other. I don't know if there is indeed such a thing as an internal relation; but, at least, that's the idea.

### 3.8 The Puzzle about the Statue and the Clay

There is a philosophical puzzle related to the issue of essential properties. Suppose Goliath (named after the legendary warrior who lost to David) is a statue made out of a single lump of clay we will call Lump1. The top half and the bottom half of Goliath are created separately and then put together to make Goliath; so Lump1, being that specific single lump, was created at the same time as Goliath. Later, Goliath is crushed to smithereens; so Lump1 is destroyed at the same time as Goliath. Lump1 starts and stops existing at the same time as Goliath does. Furthermore, it has exactly the same shape, size, location, mass, temperature, etc., as Goliath at each point of their existence. Lump1 is indistinguishable from Goliath in its entire life. Is Lump1 the same thing as Goliath? Are they numerically identical?

There is good reason to say 'no'. An essential property of Goliath is *being a statue* while an essential property of Lump1 is *being that single lump of clay*. We can think of a possible world in which the same Goliath exists but not Lump1, for instance, Goliath loses one of its hands. We can also think of a possible world in which the same Lump1 exists but not Goliath, for instance, Lump1 is flattened. Lump1 is indistinguishable from Goliath in the actual world, but distinguishable in those possible worlds. Even though they totally coincide with each other in four-dimensional spacetime in the actual world, they do not coincide in another 'dimension' – the modal dimension.

Put slightly differently, they have different modal properties. Therefore, even though Lump1 makes up or *constitutes* Goliath, it is not the same thing as Goliath. They are numerically distinct. Generalizing, two things can be in the same place at the same time throughout their lives (in the actual world or some other, but not all, worlds). This view is called *the constitution view*. Its slogan is ‘Constitution is not identity’.

Some advocates of the constitution view introduce the concept of *contingent identity* (in analogy with *temporary identity*, discussed in Section 1.5) and maintain that Goliath and Lump1 are *contingently identical* with each other in the actual world but not *necessarily identical*. If *a* and *b* are numerically identical, they are in fact one and the same thing; so they are necessarily identical, identical in all possible worlds. Numerical identity entails *necessary identity*. In contrast, if *a* and *b* are only contingently identical in the actual world, then they are in fact two distinct things and distinguishable in some other possible worlds. So Goliath and Lump1 may be contingently identical in the actual world but not numerically identical, according to this view.

Some philosophers disagree, however, and argue that Lump1 and Goliath must be one and the same thing. They maintain, among other things, that modal properties of things must have non-modal basis in the actual world. For instance, a lump of salt has the property of *being soluble in water*, i.e., it has the modal property of *being able to be dissolved in water*. But it has this modal property because of the chemical structure of salt in the actual world. Since there is no actual, non-modal difference between Lump1 and Goliath, they cannot have different modal properties; so, they argue, they cannot be distinct. To this one cannot simply reply that Goliath has the property *being a statue* that Lump1 does not have; for it is equally puzzling that Goliath can have this property that Lump1 does not have while having the same shape, size, temperature, etc., as Lump1. This problem of finding non-modal basis in the actual world for modal properties is called *the problem of grounding*. I will say a little bit more about this problem in Section 5.9.

### 3.9 De Re and De Dicto Modality

Aristotle was the most influential philosopher in antiquity. Could the most influential philosopher in antiquity have been a non-philosopher? Like some of the questions before, the answer can be ‘yes’, and it can be ‘no’, and each is correct in some sense. Again, the challenge is how to clarify those different senses.

Those who say ‘no’ would say something like this in their defense: It is self-contradictory to say ‘The most influential philosopher in antiquity is a non-philosopher’; this claim is not true in any possible world (assuming, as we do throughout this book, that we don’t change the meanings of any of the words involved).

Those who say ‘yes’ would say something like this in their defense: The most influential philosopher in antiquity is Aristotle; but Aristotle could have been a non-philosopher, not taking up philosophy in his entire life. There are possible worlds in which Aristotle exists but is not a philosopher.

When we say the latter, we are talking about *de re* modality, and when we say the former, we are talking about *de dicto* modality. ‘De re’ and ‘de dicto’ are Latin words that mean ‘of things’ and ‘of words’, respectively. In Latin, ‘res’ means ‘thing’ and ‘dicto’, as in ‘dictionary’, ‘dictation’, and ‘contradiction’, means ‘word’. When we say that the most influential philosopher in antiquity, namely Aristotle, could have been a non-philosopher, we are talking about the thing, the man, Aristotle, and saying, *of that thing*, that it could have been a non-philosopher. It really does not matter how we refer to him; instead of ‘the most influential philosopher in antiquity’, we can refer to him as ‘the teacher of Alexander the Great’, ‘the most famous disciple of Plato’, ‘the author of *Metaphysics*’, or simply ‘Aristotle’. Of course, Aristotle could have been a non-philosopher. What is important is to refer to Aristotle, regardless of how, and say *of him* that he could have been a non-philosopher. We are talking about the modality of that thing, that man, Aristotle. This is *de re* modality.

In contrast, when we say that the most influential philosopher in antiquity could not have been a non-philosopher, we are talking about the expression ‘The most influential philosopher in antiquity is a non-philosopher’, and saying that it cannot be true. This is *de dicto* modality. Here we cannot simply replace ‘the most influential philosopher in antiquity’ with co-referential but non-synonymous expressions such as ‘the teacher of Alexander the Great’ and ‘the most famous disciple of Plato’.

It is very important to clearly distinguish *de re* and *de dicto* modality when we talk about modality. One way to grammatically express *de re* modality is to insert an ‘of (or about) such and such’ phrase into ‘It is necessary/possible that ...’. For *de dicto* modality, we can quote the relevant sentence and add ‘is true’ at the end. So,

- It is necessary that the most influential philosopher in antiquity is a philosopher

will turn into

- It is necessary of the most influential philosopher in antiquity that s/he is a philosopher – *de re*;
- It is necessary that ‘The most influential philosopher in antiquity is a philosopher’ is true – *de dicto*.

For the second, keep in mind that we are not supposed to change the meanings of the expressions in the quotation.



### 3.10 ‘The Trinity Thesis’

This chapter began with a discussion of four distinctions in truths: logical vs non-logical truths, analytic vs synthetic truths, a priori vs a posteriori truths, and necessary vs contingent truths; but the discussion moved gradually to that about modality. It is time, however, to compare the distinctions, particularly the last three distinctions.

Before we begin, let us recall one important fact about necessary truth: there is exactly one (or *the*) necessary truth.

Now, historically philosophers by and large did not make any distinction between analytic, a priori, and necessary truths (with a few exceptions, as we will see). It is in fact Saul Kripke who pointed out that these concepts belong to different categories – the semantic, the epistemic, and the metaphysical, respectively. But even though they are concepts of different kinds, it is still possible that they *extensionally coincide*; that is, it is still possible that all and only necessary truths are a priori truths, and they and only they are analytic truths (more precisely, analytically true propositions). Taking into account the fact that there is exactly one necessary truth, i.e., *the* necessary truth, I will state this position as follows:

- ‘The Trinity Thesis’  
The necessary truth is the same thing as *the* a priori truth, which is the same thing as *the* analytically true proposition. That is (Figure 3.1):

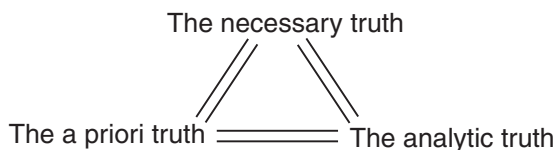


Figure 3.1 The Trinity Thesis.

Warning: unlike most other key terms in this book, which are well-known, commonly used terms, the term ‘the Trinity Thesis’ is my own invention.<sup>4</sup> Hence the scare quotes. I use the term because I think it’s very appropriate. I hope it will catch on. But don’t use it as if it were a common currency.

There is *prima facie* reason to think that the necessary truth is all and the only a priori truth. To obtain a priori knowledge, we cannot use experience to know what the world is like; we have to run all possibilities in our head. But we cannot know a priori a proposition which *just happens* to be true in the actual world, for that proposition would be false or valueless in some other worlds, and without experience we cannot know which of those worlds is the actual world. So the only a priori proposition we can know

to be true in the actual world must be the proposition true in all possible worlds, i.e., the necessarily true proposition.

As I suggested earlier, there is reason to think that all analytically true sentences express propositions that can be known a priori. That's because if a sentence is an analytic truth, we can simply analyze in our head the meanings of the words involved and determine that the sentence is true. In fact, there is reason to think that the other way around is also true, i.e., that the analytically true sentences express only true propositions that can be known a priori. If we cannot use experience to synthesize facts around us, the only way to obtain truths seems to be to analyze the meanings of sentences.

These considerations lead us to the Trinity Thesis. It is, again, a very plausible thesis. In fact, even though it is Kripke's great contribution to point out that necessity, apriority, and analyticity are concepts that belong to different categories, the reason why philosophers in the past tended to conflate them might be that they believed in something like the Trinity Thesis. If the necessary truth = the a priori truth = the analytic truth, it does not seem so significant that they belong to different categories.

Immanuel Kant and Kripke, however, argued against the Trinity Thesis; in particular, Kant argued that there were synthetic a priori truths, and Kripke argued that there were necessary a posteriori truths. In the next two sections we will consider their arguments. We will see that they both fail.

### **3.1.1 Kant's Synthetic A Priori**

It was the German philosopher Immanuel Kant (1783) who made famous the distinctions between the analytic and the synthetic and the a priori and the a posteriori. Kant argued that, contrary to what one may think, there is such a thing as synthetic a priori truths, i.e., truths which express propositions you can learn a priori, which nonetheless say something substantive about the world. Kant maintained, among other things, that mathematical truths, which can be known a priori, are nonetheless synthetic, telling us about the structure of the actual space and time. For instance, the sum of the inner angles of every triangle is  $180^\circ$ .<sup>5</sup> We can derive this proposition by reason alone from Euclid's axioms, which are themselves a priori truths. At the same time, it is a synthetic truth, telling us about the 'shape' of actual space.

Kant then asked: how are these synthetic a priori truths possible? His surprising answer was that they are possible because space and time are not objective features of the world existing independently of us; instead, they are the framework of our mind. Just as we will see everything dark if we are wearing dark-colored sunglasses, we see everything in spacetime because we are wearing spacetime sunglasses. We can know features of these spacetime sunglasses a priori because they are part of us. Unlike the dark-colored sunglasses, however, we can never take off the spacetime sunglasses;

so we can never know what the real world is like, Kant argued. He called the things-as-we-experience-them *Phenomena* and the things-in-themselves *Noumena*, and contended that even though we never know what the world of Noumena is like, it must exist behind the world of Phenomena.

Interesting as Kant's argument may be, it is now generally considered flawed. Kant is often accused of being guilty of conflating *pure* and *applied* mathematics.

The discovery of non-Euclidean geometries, which took place in the mid to late nineteenth century shortly after Kant's death, has revealed his mistakes. The story culminating in the discovery is one of the most fascinating intellectual dramas in history, spanning over twenty centuries.

Euclid wrote thirteen Books of *Elements* around 300 B.C. It was possibly the second most well-read book in Western history only after the Bible. Until the twentieth century, most educated people studied it in its original form at school. It derives many complex theorems in geometry in a rigorous, step-by-step fashion from a few definitions, postulates, and axioms. You may know at least that Pythagoras's Theorem, i.e., that the square of the hypotenuse of a right triangle equals the sum of the squares of its two sides, is proved there (Book 1, Proposition 47). Euclid proves this theorem and many others only from five properly geometrical postulates and nine general axioms. The theorems were regarded as universal truths and their proofs as an exemplar of how to obtain a priori knowledge deductively.

The famous five postulates of *Elements* are as follows:

- (1) There is a straight line between any two points.
- (2) Any straight line can be extended indefinitely.
- (3) For any line segment starting at any point, there is a circle with that point as center and that line segment as radius.
- (4) All right angles are equal to one another.
- (5) If a straight line falls on two straight lines in such a manner that the interior angles on the same side are together less than two right angles, then the straight lines, if produced indefinitely, meet on that side (Parallel Postulate; Figure 3.2).

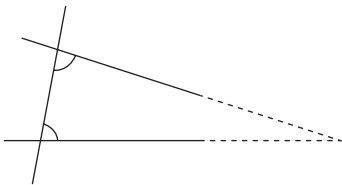


Figure 3.2 Parallel Postulate.

However, throughout history many people had trouble with the Fifth Postulate (5), the Parallel Postulate. It is much more complicated and much less self-evident than the other four. Thus, many mathematicians tried to derive the Fifth Postulate from the other four. It became known that given the other four postulates, (5) is equivalent to (6) below; that is, (6) is derivable from (1) to (5), and (5) is derivable from (1) to (4) and (6).

- (6) There is at most one line that can be drawn parallel to another given line through an external point (Playfair's Axiom).

Here parallel lines are lines which will never meet no matter how much they are extended. Because (6) is equivalent with (5) but easier to deal with, people tried to prove (6) from (1) to (4). (5) (thus, also (6)) is also equivalent to the following postulate:

- (7) The sum of the angles in every triangle is  $180^\circ$  (Triangle Axiom).

I will abbreviate this statement as ' $\Delta = 180^\circ$ ' in what follows.

For over two thousand years mathematicians tried to prove (6) from (1) to (4) but could not do it, until the 1830s when the Russian mathematician Nikolai Lobachevsky and the Hungarian mathematician János Bolyai independently published works on what is now called *hyperbolic geometry*, which assumes (1) to (4) but the denial of (6). In hyperbolic geometry, for any line and any point external to it, infinitely many lines can be drawn through the point parallel to the line, and  $\Delta < 180^\circ$ . A little later Bernhard Riemann presented to the world what is now called *elliptic geometry*, according to which there is no external line parallel to a given line, and  $\Delta > 180^\circ$ .

How are those geometries possible? For the sake of simplicity, let's suppose we live in two-dimensional space instead of three. In the case of hyperbolic geometry, imagine that we are living in the disk inside a big circle, where the boundary – the circle itself – is not part of the world (Figure 3.3).

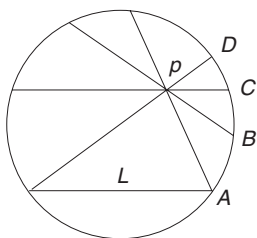


Figure 3.3 A model of two-dimensional hyperbolic geometry.

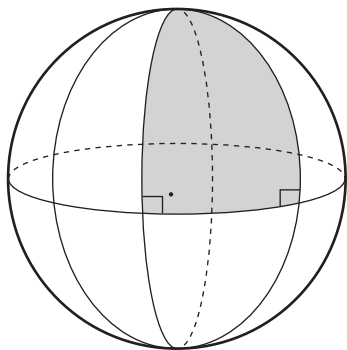


Figure 3.4 A model of two-dimensional elliptic geometry.

Imagine that things get shorter and smaller, and their motion slower, as they move from the center to the boundary of this world; so if you travel from the center toward the boundary, you will never reach the boundary. In that respect, this world has infinite space.  $L$  is a straight line. Take any point  $p$  outside of  $L$ . Then  $A, B, C, D \dots$  are all straight lines through  $p$  parallel to  $L$  because they don't intersect with  $L$  in this world. ( $A$  and  $D$  meet  $L$  only on the circle, which, however, is not part of the world.) Clearly there are infinitely many straight lines parallel to  $L$ .

For two-dimensional elliptic geometry, imagine that our world is the surface of a sphere (Figure 3.4). Here straight lines are 'great circles' on the surface, i.e., circles whose centers are the center of the sphere. Then there is no straight line outside of and parallel to any given straight line, for any two great circles intersect somewhere. The sum of the inner angles of a triangle, like the shaded one, will be more than  $180^\circ$  (i.e.,  $\Delta > 180^\circ$ ).

While it is true that the claim  $\Delta = 180^\circ$  is true in Euclidean geometry, a competing claim  $\Delta < 180^\circ$  is true in hyperbolic geometry, and another competing claim  $\Delta > 180^\circ$  is true in elliptic geometry. And which geometry captures the actual 'shape' of real, physical, space, and whether  $\Delta = 180^\circ$  or  $< 180^\circ$  or  $> 180^\circ$  in real space, are purely empirical questions you cannot answer a priori. The famous mathematician Gauss tried to answer the question by taking actual measurements between the peaks of three German mountains, and Lobachevsky made a similar attempt, even though their results were inconclusive. In fact, Einstein's General Relativity Theory has shown that, in real space,  $\Delta > 180^\circ$ .<sup>6</sup> What Kant would have called an a priori truth, i.e.,  $\Delta = 180^\circ$ , has turned out to be not even true!

To summarize, we must distinguish the truths of pure math from the truths of applied math. The truths of pure math, such as

- $\Delta = 180^\circ$  in Euclidean geometry (i.e., the axioms of Euclidean geometry entail  $\Delta = 180^\circ$ );
- $\Delta < 180^\circ$  in hyperbolic geometry (i.e., ...);
- $\Delta > 180^\circ$  in elliptic geometry (i.e., ...);

are a priori but analytic truths. The truths of applied math, such as

- $\Delta > 180^\circ$  in real space,

are synthetic but a posteriori truths. Contrary to Kant, there is no room we can find for the synthetic a priori.

### 3.12 Kripke's Necessary A Posteriori

Kripke's argument for the existence of necessary a posteriori truths goes like this: Recall the story of Hesperus and Phosphorus. At one point of history, we made the astronomical discovery that Hesperus is identical with Phosphorus (i.e., that  $\text{Hesperus} = \text{Phosphorus}$ ). Thus, that  $\text{Hesperus} = \text{Phosphorus}$  is a posteriori knowledge. Recall (Section 2.8), however, that the proper names 'Hesperus' and 'Phosphorus' are rigid designators denoting one and the same thing, the planet Venus, in the actual world. So, as we confirmed, they have the same intension. Hence, just as it is necessary that  $\text{Hesperus} = \text{Hesperus}$ , it is necessary that  $\text{Hesperus} = \text{Phosphorus}$ . Therefore, that  $\text{Hesperus} = \text{Phosphorus}$  is a necessary but a posteriori truth.

Kripke used the same argument to show that many statements of theoretical identity, such as that  $\text{water} = \text{H}_2\text{O}$  and that  $\text{heat} = \text{mean molecular kinetic energy}$ , are statements of necessary identity even though we discover them a posteriori. This argument has convinced many people of the existence of necessary a posteriori truths. There are in fact still numerous philosophers who believe and say openly that Kripke has shown that there are necessary a posteriori truths.

Among the experts on this subject, however, it is widely acknowledged that Kripke did not succeed in showing the existence of the necessary a posteriori. For instance, Robert Stalnaker, another prime architect of possible worlds semantics, responded to Kripke's argument as follows (Stalnaker 1976, with my modifications):

'Hesperus' and 'Phosphorus' do indeed have the same intension; there is no difference in the meanings of 'Hesperus' and 'Phosphorus'. So the proposition that if Hesperus exists,  $\text{Hesperus} = \text{Phosphorus}$  is the same proposition as that if Hesperus exists,  $\text{Hesperus} = \text{Hesperus}$ , true in all possible worlds. Since we knew that if Hesperus exists,  $\text{Hesperus} = \text{Hesperus}$  before the astronomical

discovery, we knew that if Hesperus exists, Hesperus = Phosphorus before the discovery. Thus, that if Hesperus exists, Hesperus = Phosphorus, though it is in fact the necessary truth, is not a posteriori knowledge. Of course, we did not know before the discovery that this proposition, i.e., that if Hesperus exists, Hesperus = Hesperus, or that if Hesperus exists, Hesperus = Phosphorus, could be expressed by the sentence 'If Hesperus exists, Hesperus = Phosphorus'. We only knew that it could be expressed by the sentence 'If Hesperus exists, Hesperus = Hesperus'. What we came to know after the astronomical discovery, what the discovery showed us, is that the same proposition could be expressed by another sentence, 'If Hesperus exists, Hesperus = Phosphorus'.

To summarize, before the discovery we knew:

- (a) that if Hesperus exists, Hesperus = Hesperus (or Phosphorus);
- (b) that the sentence 'If Hesperus exists, Hesperus = Hesperus' expresses proposition (a) that if Hesperus exists, Hesperus = Hesperus (or Phosphorus).

After the discovery we came to know:

- (c) that the sentence 'If Hesperus exists, Hesperus = Phosphorus' expresses proposition (a) that if Hesperus exists, Hesperus = Hesperus (or Phosphorus).

(a) is the necessary truth, but (b) and (c) are contingent facts about our use of language. So, at the point of the discovery we did not come to believe a necessarily true proposition, contrary to Kripke's claim, Stalnaker argued.

This solves – or at least gives one solution to – Frege's Puzzle. Frege's Puzzle, recall (Section 2.4), was that if 'Hesperus' and 'Phosphorus' don't have different intensions (or senses), how can you explain the cognitive difference between '(If Hesperus exists) Hesperus = Hesperus' and '(If Hesperus exists) Hesperus = Phosphorus'? The answer being given here is that even though 'Hesperus' and 'Phosphorus' do not have different intensions and, thus, 'If Hesperus exists, Hesperus = Hesperus' and 'If Hesperus exists, Hesperus = Phosphorus' express one and the same proposition, the necessarily true proposition, the fact that the sentence 'If Hesperus exists, Hesperus = Hesperus' expresses that proposition is obvious and was known before the astronomical discovery, whereas the fact that the sentence 'If Hesperus exists, Hesperus = Phosphorus' expresses that proposition is not obvious and can be known only at the point of the astronomical discovery. That explains the apparent difference in the cognitive significance of 'Hesperus' and 'Phosphorus'.

We can say the same thing about statements of theoretical identity. Even before Cavendish's discovery, we knew that if water exists, water = H<sub>2</sub>O,

for we knew that if water exists, water = water. We did not know that this identity could be expressed as 'If water exists, water = H<sub>2</sub>O', too.

This theory draws a particularly interesting picture about mathematical knowledge. Mathematical objects such as sets and numbers are, unlike Venus or water, *necessary existents (or beings)*: they exist in all possible worlds. Thus, true mathematical equations express one and the same proposition, the necessary truth. Furthermore, it seems that there is no point of time in our lives that we *came to* grasp this proposition. In the case of Venus or water, we may come to know that Hesperus = Hesperus or that water = water when we come to know that Venus or water exists in the actual world; but there is no such point of time in the case of mathematical objects. So, perhaps, our knowledge of the mathematical truth is not only a priori but *innate*: we are born with that mathematical knowledge. When we prove an equation, we only find a new way of expressing something we already knew. Perhaps Leibniz was right about that.<sup>7</sup>

You might think that this is bizarre. But it is a consequence of possible worlds semantics, the semantics Kripke, among others, has developed. As I said in Section 2.9, the concept of intension in the possible worlds semantics is often criticized as being too coarse-grained, not distinguishing the apparent differences. Whether this criticism is right or not, the picture of necessary truth I have just depicted is a consequence of the concept of intension developed in possible worlds semantics. Thus, we can conclude that Kripke did not succeed in showing that there are necessary a posteriori truths.<sup>8</sup>

To summarize the last two sections, neither Kant nor Kripke has succeeded in breaking up the Trinity. It is still possible that the content of a priori knowledge is the necessary truth, and that all and only analytically true sentences express this unique proposition.

Some might say that the last part of the Trinity, i.e., that only analytically true sentences express the necessary truth, is obviously incorrect. Think of the sentence 'If Hesperus exists, then Hesperus = Phosphorus'; this sentence is a synthetically true sentence but expresses the necessary truth, they might say. I hope you have mastered the subject well enough to give the correct reply to this claim, which is (fanfare!) to deny that the sentence is synthetically true. It is *analytically* true in the current framework. For the meaning (intension) of 'Phosphorus' is no different from that of 'Hesperus', and, thus, the meaning of 'If Hesperus exists, Hesperus = Phosphorus' is no different from that of 'If Hesperus exists, then Hesperus = Hesperus'; but this sentence is analytically true and expresses the necessary truth. Therefore, 'If Hesperus exists, then Hesperus = Phosphorus' is also analytically true and expresses the necessary truth. The claim that 'If Hesperus exists, then Hesperus = Phosphorus' is analytic, admittedly, is very counterintuitive. But it is, again, a consequence of the coarse-grainedness of possible worlds semantics we have adopted. So until and unless we replace it with something better, the above consideration does not break the Trinity.



Some philosophers have suggested revising or replacing the concepts of proposition, and intension in general, defined in possible worlds semantics. Some, such as Jackson (1998) and Chalmers (2004, 2006), have proposed *two-dimensional semantics*, which distinguishes two kinds of intension, the *primary* and the *secondary intension*. According to this theory, 'Hesperus' and 'Phosphorus' have the same secondary intension but different primary intensions. Others (Cresswell 1985) have introduced the concept of *structured propositions*. On the conception of proposition we discussed, a proposition is just one thing, one function, with no internal structure. In contrast, a structured proposition is related to certain individual concepts and properties just as a sentence is related to its constituent singular terms and predicates. Thus, 'If it rains, then it rains' and ' $2+2=4$ ' express different structured propositions. It is still a matter of debate whether these new concepts are in fact better tools than intension and proposition defined in possible worlds semantics, but we won't get into that discussion in this book. It is much more important for the readers of this book to have a firm grasp of the basic ideas of possible worlds semantics.

### 3.13 Counterfactual Conditionals

In the rest of this chapter, I would like to touch on a few topics related to modality and possible worlds semantics: counterfactual conditionals, causation, and three non-metaphysical modalities – epistemic, deontic, and temporal modality. This section deals with the first topic: counterfactual conditionals.

There are three kinds of conditional we encounter when we do logic or philosophy: the material conditional, the indicative conditional, and the subjunctive (or counterfactual) conditional. To repeat, the conditional  $\rightarrow$  defined by the truth table given in Section 1.3 is called *the material conditional*. The material conditional, used in logic, is not the same as two kinds of conditionals contained in English, *the indicative conditional* and *the subjunctive (or counterfactual) conditional*. The difference between them can be explained as follows:

Suppose today is Monday. Compare:

- (a) (Today is Sunday)  $\rightarrow$  (tomorrow is Wednesday) – material conditional;
- (b) If today is Sunday, tomorrow is Wednesday – indicative conditional;
- (c) If today were (or was) Sunday, tomorrow would be Wednesday – subjunctive conditional.

(a) is true because the antecedent 'today is Sunday' is false, and a material conditional is true regardless of its consequent if its antecedent is false. (Check the truth table if you are not sure about this.) If I don't know what day of the week today is, I may assert (b). If I know that today is Monday,

I may assert (c). Both conditionals are false, however, whereas (a) is true. In general, neither of the two conditionals in English, the indicative conditional and the subjunctive conditional, is the same as the material conditional, the conditional used in propositional logic.

The difference between the subjunctive conditional and the indicative conditional is often illustrated with the following example involving the JFK assassination:

- (d) If Oswald had not shot Kennedy, someone else would have – subjunctive conditional;
- (e) If Oswald did not shoot Kennedy, someone else did – indicative conditional.

(d) is false given that the Kennedy assassination was Lee Harvey Oswald's solo act (the conspiracy theory is simply false). In contrast, (e), if it was asserted before Oswald's guilt was confirmed, seems to have been true since someone did shoot Kennedy. In general, the subjunctive conditional and the indicative conditional are of different kinds.

So, the subjunctive conditional is different from the material or the indicative conditional. What linguists call the subjunctive conditional is called *the counterfactual conditional* by philosophers. Generally (with some exceptions), a counterfactual conditional is a conditional sentence whose antecedent (the if-part) postulates a counterfactual (i.e., contrary-to-fact) state of affairs. Schematically, a typical counterfactual conditional can be expressed in one of these forms:

- If it were the case that  $P$ , then it would be the case that  $Q$ ;
- If it had been the case that  $P$ , then it would have been the case that  $Q$ .

We will symbolize these counterfactuals as ' $P > Q$ '.<sup>9</sup>

The following are a few examples of counterfactual conditional:

- (f) If the moon were made of cheese, pigs would fly.
- (g) If today were Thanksgiving Day, tomorrow would be Black Friday. (Assuming that you have better things to do than reading this on Thanksgiving Day.)
- (h) If Steve Bartmann had not touched the foul ball, Moses Alou would have caught it, and the Chicago Cubs would have won the 2003 National League Championship Series.
- (i) If FBI director James Comey had not reopened the email investigation in October, Hillary Clinton would have won the 2016 Presidential election.

Since counterfactual conditionals are different from material or indicative conditionals, they must be given their own distinct interpretation, interpretation different from those given to the other two kinds. This is where possible worlds seem to give us much help. For instance, in his *possible worlds analysis of counterfactuals*, David Lewis (1973a; also Stalnaker 1968) introduced the concept of *closeness* among the possible worlds and gave an analysis of  $P > Q$  in terms of it:

- Possible worlds analysis of counterfactuals
  - (1)  $P > Q$  is vacuously true iff there is no possible world in which  $P$  is true;
  - (2)  $P > Q$  is non-vacuously true iff among the possible worlds in which  $P$  is true (or ‘the  $P$ -worlds’), some possible world in which  $Q$  is true (some  $Q$ -world) is *closer* to the actual world than any possible world in which  $Q$  is not true (any non- $Q$ -world);
  - (3) Otherwise,  $P > Q$  is false.

Closeness is a tricky concept, so I won’t try to clarify here. Suffice it to say, for your initial grasp of the theory, that it’s overall similarity between worlds.

According to this theory, (f) above is probably false because among the worlds in which the moon is made of cheese, some worlds in which pigs don’t fly are probably at least as close to the actual world as any world in which pigs fly. On the other hand, (g) is most likely true because among the worlds in which today is Thanksgiving Day, many worlds in which tomorrow is Black Friday is closer to the actual world than any world in which tomorrow is not Black Friday. The first part of (h) seems true, given that Alou was just under the ball when Bartman touched it; among the possible worlds in which Bartman did not touch the ball, there seem to be many possible worlds closer to the actual world in which Alou caught the ball than any worlds in which he did not catch it. The second part about the Cubs’ destiny is more questionable. (Recall that there was the rest of the sixth game and all of the seventh game.) The truth and falsity of (i) are quite a matter of controversy. However, while the present theory may not give us a definitive answer to the question whether the relevant counterfactual conditional is true, it does tell us what the question *means* and how, theoretically, we can determine the answer. Namely, (i) says that among the worlds in which James Comey did not reopen the email investigation, some worlds in which Hillary Clinton won the election are closer to the actual world than any worlds in which Clinton still lost to Donald Trump. The possible worlds analysis of counterfactuals is helpful at least in that respect.

One problem with Lewis’s counterfactual analysis concerns condition (1) about the vacuous cases. (1) states that every counterfactual conditional

of form  $P > Q$  is true if its antecedent  $P$  is impossible. That does not seem correct, however. For instance, compare the following sentences:

- (j) (For any sentence  $A$ ) if  $A$  were both true and false, then  $A$  would be true (or false);
- (k) (For any sentence  $A$ ) if  $A$  were both true and false, then  $A$  would be neither true nor false.

(j) seems true but (k) seems false. So condition (1) seems implausible.

### 3.14 Causation

The next topic, which is closely related to the last, is causation. David Hume's theory of causation is well known; it is called *the constant conjunction (or regularity) theory of causation*. Hume argued that when you see one event, say  $E$  (the yellow billiard ball's moving) caused by another event, say  $C$  (the yellow billiard ball's being hit by the red billiard ball), you don't observe the *necessary connection* between the cause  $C$  and the effect  $E$ ; you don't observe  $C$ 's *forcing*  $E$ . All you can observe is that every  $C$ -like event has always been followed by an  $E$ -like event, i.e., a constant conjunction (or regularity) between  $C$ -like events and  $E$ -like events. However, if you observe many instances of the conjunction, your mind becomes disposed to connect  $C$ -like events with  $E$ -like events, and this disposition of the mind is the source of our *feeling* of the necessary connection between  $C$  and  $E$ . But the necessary connection between  $C$  and  $E$  does not exist in reality. That's basically Hume's view.

However, Hume at one point made the following notorious statement:

We may define a cause to be *an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second*. Or, in other words, *where, if the first object had not been, the second never had existed*.

(Hume 1748, Section VII, Part II)

This statement is notorious because what is said before and after "Or, in other words" is not the same thing at all. While the first statement is the expression of the constant conjunction theory of causation, which denies the necessary connection between the cause and the effect, the second statement – which, notice, is a counterfactual conditional – seems to state exactly that kind of necessary connection between the cause and the effect.

The contemporary counterfactual theories of causation take a cue from the second part of the above quotation. In particular, Lewis (1973b) made use of his theory of counterfactuals we discussed in the last section and offered the following *counterfactual analysis of causation*:

- Counterfactual analysis of causation  
Event *C* caused event *E* iff
  - (a) *C* occurred, and *E* followed; and
  - (b) if *C* had not occurred, *E* would not have occurred.

For instance, those who say that James Comey's reopening of the email investigation caused Clinton's defeat are saying that (a) the reopening occurred and the defeat followed; and that (b) if the reopening had not occurred, the defeat would not have occurred (i.e., Clinton would have won), where the counterfactual conditional (b) can be interpreted in terms of possible worlds as we saw in the last section.

Just like the possible worlds analysis of counterfactuals, the counterfactual analysis of causation does not tell us in a straightforward fashion when something caused something else; instead, it only gives us an analysis of what causation ought to be. It helps us, however, to sharpen our sense about causation and contributes greatly to the philosophical discussion about it.

### 3.15 Epistemic and Deontic Modality

Finally, as I said in Section 3.6, I would like to say a little about kinds of modality other than metaphysical modality, which was a major focus of attention in this chapter. In particular, I would like to say a little about *epistemic* and *deontic modality* in this section and *temporal modality* in the next section. 'Deon' in Greek means 'duty'. You may know that deontology in ethics is the duty-based moral theory. So deontic modality is the modality of moral obligation and permissibility.

As I said in Section 3.6, 'must' and 'can' usually express necessity and possibility; but the necessity and possibility in question may not be those of metaphysical modality. If I say 'The keys can be in my car or on my desk; but they must be in one of those two places', I am not saying that in every metaphysically possible world, the keys are either in my car or on my desk. If I say 'We must treat everybody equally; but we still can give Christmas presents to our children', I am not saying that in every metaphysically possible world, we treat everybody equally. The former 'must' and 'can' mean epistemic certainty and possibility, and the latter 'must' and 'can' mean moral obligation and permissibility.

The logics of epistemic modality and deontic modality are called *epistemic logic* and *deontic logic*, respectively. Even though epistemic and deontic modality are different from metaphysical modality, we still can devise possible worlds semantics for them. The key here is to introduce the concept of *accessibility relation* among worlds. In the case of metaphysical modality, whether proposition *P* is necessary or not in the actual world – or in *any* world, for that matter – is determined by whether *P* is true in *absolutely all* possible worlds or not, and whether *P* is possible or not in the actual world

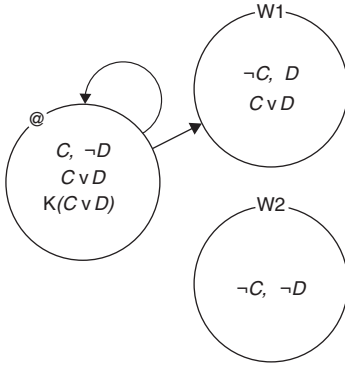


Figure 3.5 Epistemic possible worlds.

or in any other world is determined by whether  $P$  is true in at least one of *absolutely all* possible worlds or not. In contrast, in the case of epistemic or deontic modality, each world has access only to a limited group of possible worlds, and the epistemic or deontic necessity and possibility in that world are determined with respect only to the worlds it has access to. *The accessibility relation*  $R(v, w)$ , i.e.,  $v$  has access to  $w$ , is determined among the worlds.<sup>10</sup>

For instance, let's define:

- $C$ : The keys are in my car;
- $D$ : The keys are on my desk.

Let's suppose that the keys are in fact in my car and not on my desk (i.e.,  $C$  and  $\neg D$ ) in the actual world  $@$  (Figure 3.5).

However, I can conceive of an alternative scenario (i.e., an epistemic possible world  $W1$ ) in which  $\neg C$  and  $D$ . That means that the world I live in has epistemic access to that world as well as the world  $@$  in which  $C$  and  $\neg D$ . On the other hand, it does not have access to world  $W2$  in which  $\neg C$  and  $\neg D$ , for such a scenario is inconceivable to me. That is,  $R(@, @)$ ,  $R(@, W1)$ , but not  $R(@, W2)$ . Now, let's introduce the operator:

- $K(X)$ : It is known to me that  $X$ .

This is a necessity operator of sort:

- $K(P)$  is true in  $w$  iff  $P$  is true in all the possible worlds accessible from  $w$ .

If the right-hand side holds, not- $P$  is inconceivable to me in  $w$ . Since  $C \vee D$  in both  $@$  and  $W1$ , in all the worlds accessible from  $@$ ,  $K(C \vee D)$ , that is,

I know that  $C \vee D$ . (Needless to say, I am simplifying here, for in fact there must be numerous possible worlds the actual world @ has access to, each being slightly or largely different from one another.)

Note that, just as  $\Box P$  implies  $P$  in metaphysical modality,  $K(P)$  implies  $P$  in epistemic modality. This is in harmony with our concept of knowledge according to which I know that  $P$  implies that  $P$  is true. (We will talk more about the concept of knowledge later in Section 5.10.) For this implication relation to hold generally, every world must have access to itself: the accessibility relation must be reflexive (i.e.,  $R(w, w)$  for any world  $w$ ). For if it is not, then some world  $w$  may contain both  $\neg P$  and  $\Box P$ , i.e.,  $P$  in all the worlds accessible from  $w$ .

One recent debate pertaining to epistemic modality concerns the question whether the following principle should hold:

- The KK Principle  
 For any sentence  $P$ ,  
 $K(P) \rightarrow KK(P)$  (i.e., if I know that  $P$ , then I know that I know that  $P$ )  
 holds in any possible world  $w$ .

Note that the corresponding metaphysical conditional  $\Box P \rightarrow \Box \Box P$  (i.e., if it is necessary that  $P$ , then it is necessary that it is necessary that  $P$ ) holds in any possible world  $w$ , as the following argument shows: Take any world  $w$ , and suppose  $\Box P$  holds in  $w$ . Then  $P$  holds in all possible worlds. So  $\Box P$  holds in all possible worlds. Thus,  $\Box \Box P$  holds in  $w$ . Therefore,  $\Box P \rightarrow \Box \Box P$  holds in  $w$ . QED. However, the analogous argument for  $K$  may not go through if the accessibility relation is restricted. The question is whether the KK Principle is a plausible principle we should embrace for epistemic modality. The KK Principle will hold iff the accessibility relation for epistemic modality is transitive (i.e., if  $R(u, v)$  and  $R(v, w)$ , then  $R(u, w)$  for any worlds  $u, v$ , and  $w$ ). (Why? Think about it.)<sup>11</sup> So the question is: Should the accessibility relation for epistemic modality be transitive? We will come back to this question in Chapter 5 when we discuss the internalism–externalism debate in epistemology.

Another important issue pertaining to epistemic modality concerns the maximality and consistency of possible worlds, assumed since when possible worlds are introduced in Section 2.5 until now. In the case of metaphysical (or, more precisely, metaphysically accessible) possible worlds, it is reasonable to think that each possible world is just like the actual world and is thus maximal and consistent; that is, for any proposition  $P$ , either  $P$  is true or not- $P$  is true in each world, and not both  $P$  is true and not- $P$  is true in each world. However, it is much less reasonable to think that these two conditions hold also for epistemic (or epistemically accessible) possible worlds. Why should a world conceivable to me be maximal and consistent?

Some philosophers are particularly skeptical about the consistency of worlds and have come to believe in the existence of *impossible possible worlds* (or *impossible worlds*).<sup>12</sup> Apparently, unlike God, we are not logically omniscient; we do not always believe a logical consequence of what we believe because we don't realize that that's the consequence. So, according to those philosophers, it is reasonable to believe in impossible possible worlds in which  $P$  and not- $P$  are both true. That we have access to such worlds does not mean that we can combine  $P$  and not- $P$  immediately and derive anything and everything. Perhaps there may even be an impossible world in which Hesperus is the planet Venus but Phosphorus is not the planet Venus.

Impossible worlds can be used for other purposes. For instance, at the end of Section 3.13 we discussed a problem with Lewis's possible worlds analysis of counterfactuals. According to that analysis,  $P > Q$  is vacuously true iff there is no possible world in which  $P$  is true; but, as we saw, this definition seems implausible if the relevant possible worlds do not include impossible worlds. But if impossible worlds are included, in particular an impossible world in which  $A$  is both true and false, then counterfactual conditional (j), if  $A$  were both true and false, then  $A$  would be true (or false), may reasonably be true while (k), if  $A$  were both true and false, then  $A$  would be neither true nor false, may reasonably be false. Needless to say, the kind of logic we have been dealing with since the early part of Chapter 1, the kind of logic we use every day, so-called *classical logic*, does not apply in impossible worlds; unlike in a possible world like the actual world, the Law of Non-Contradiction does not hold and a contradiction does not imply everything in an impossible world (cf. Exercise Question 5 in this chapter). Some *non-classical logic* must be employed instead. It is still a matter of debate whether we should accept the existence of impossible worlds or not.

But that's enough about epistemic modality and related issues. Let's move on to deontic modality; and let's consider the previous example again. Suppose:

- $T$ : We treat everybody equally.
- $G$ : We give Christmas presents to our children.

Suppose that in the actual world @, we in fact give Christmas presents to our children but don't treat everybody equally i.e.,  $\neg T$  and  $G$  (Figure 3.6).

But the actual world has access to  $W1$ , in which  $T$  and  $G$ , as well as to  $W2$ , in which  $T$  and  $\neg G$ . In this case,  $v$ 's having access to  $w$ , i.e.,  $R(v, w)$ , means that  $w$  is a morally ideal world with respect to  $v$ . Naturally, @ does not have access to @ itself, for our actual world is hardly a morally ideal world. That is,  $R(@, W1)$  and  $R(@, W2)$ , but not  $R(@, @)$ . Let's introduce two modal operators:



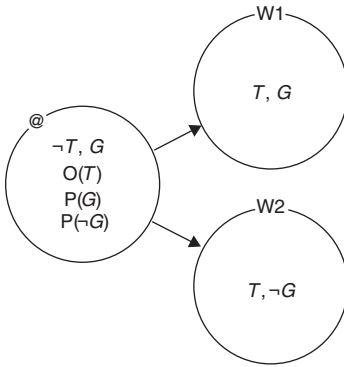


Figure 3.6 Deontic possible worlds.

- $O(X)$ :  $X$  is morally obligatory;
- $P(X)$ :  $X$  is morally permissible.

$O$  is a necessity operator of sort, whereas  $P$  is the corresponding possibility operator:

- $O(P)$  is true in  $w$  iff  $P$  is true in all the possible worlds accessible from  $w$ ;
- $P(P)$  is true in  $w$  iff  $P$  is true in at least one possible world accessible from  $w$ .

Since  $T$  is the case in all the possible worlds accessible from  $@$ ,  $O(T)$  in  $@$ .  $G$  is the case in  $W1$  whereas  $\neg G$  is the case in  $W2$ , so both  $P(G)$  and  $P(\neg G)$  in  $@$ . Treating everybody equally is a morally obligatory act whereas giving and not giving Christmas presents to our children are both morally permissible acts in the actual world.

Recall that in the case of both metaphysical and epistemic modality, necessity implies truth: if  $\Box P$ , then  $P$ , and if  $K(P)$ , then  $P$ . In contrast, deontic necessity, i.e., moral obligation, does not imply truth; even if it is morally obligatory that  $P$ ,  $P$  may not be the case, for we may not do what is morally obligatory (we may not perform our duties). From the logical point of view, this is possible because some worlds do not have access to themselves: the relevant accessibility relation is not reflexive.

As the above examples show, by putting various constraints on the accessibility relation, we can support various modal logics. In fact, there are numerous modal logics out there, supported by various accessibility relations, representing different modal concepts.

Postulating various accessibility relations to deal with various modal logics is originally Kripke's idea. Kripke introduced this idea in a rigorous form in

a mathematics paper he wrote when he was a high school student. What did *you* do when you were a high school student? (What did *I* do when I was a high school student?)

### 3.16 Temporal Modality

The last of the non-metaphysical modalities I would like to talk about is temporal modality. I introduced four-dimensionalism about spatiotemporal objects back in Section 1.5. Four-dimensionalism holds that material objects like humans are not three-dimensional spatial objects existing through time but four-dimensional objects existing in spacetime. Each four-dimensional object has temporal as well as spatial mereological parts. In Section 2.3 when I talked about mereology, I focused on spatial parts; the example given there was my body's having six spatial parts. But just as I have spatial parts, as a four-dimensional object I also have *temporal parts*, according to four-dimensionalism. Suppose, again (Section 1.5), that I was born in 1960 and will die in 2050 at age 90. Then the four-dimensional whole which is me spans over those 90 years, and is made up of many temporal parts (some of which may overlap). For instance, Ken Akiba for the first year after birth may be considered one temporal part, Ken Akiba at age 1 may be considered another, Ken Akiba at age 2 may be considered yet another temporal part, etc., of the whole 90 years of Ken Akiba. Ken Akiba in each day, each hour, each minute, or each second can be considered even smaller temporal parts.<sup>13</sup> The smallest of those temporal parts are *temporal slices*, instantaneous objects with no width in the temporal dimension. Imagine a four-dimensional salami sausage made up of infinitely thin slices cut orthogonally to the temporal axis. All temporal slices of Ken Akiba may, furthermore, be considered to

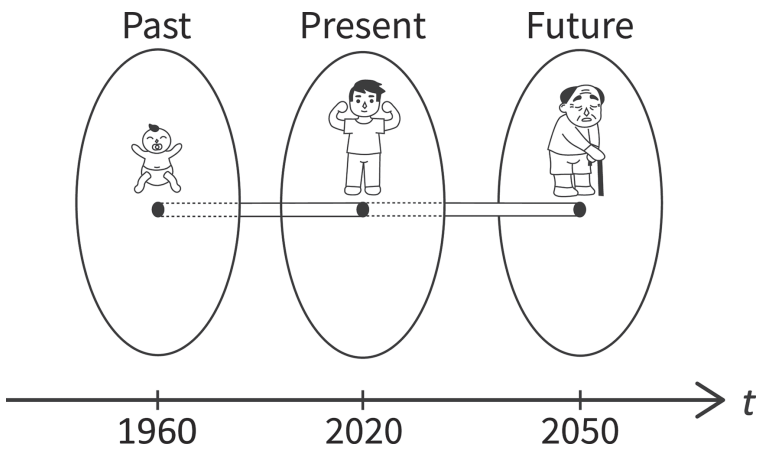


Figure 3.7 Temporal worlds.

exist in instantaneous slices of time. The whole picture looks like this (Figure 3.7). This is an instance of identity over time, discussed in Section 1.5.

Doesn't this figure remind you of something? Yes, it looks exactly like Figures 2.14, 2.17, and 2.18 of possible worlds and transworld objects. This suggests the idea that each time slice can be taken as a possible world, called a *temporal world*, and that each individual can be taken as a transworld object extending over temporal worlds. Just as Aristotle may study philosophy in some metaphysically possible worlds (including the actual world) and not study philosophy in other metaphysically possible worlds, I may study philosophy in some temporal worlds (those from 1978 to 2030, say, including the present) and not study philosophy in other temporal worlds. The only major difference between the metaphysical case and the temporal case is that in the temporal case, temporal worlds are linearly ordered from the earlier to the later and are continuous, whereas metaphysical possible worlds have no particular order.

In the deontic case, 'necessarily' and 'possibly' amount to 'obligatorily' and 'permissibly'. What do 'necessarily' and 'possibly' amount to in the temporal case? They are 'always' and 'sometimes', respectively. 'Always  $P$ ' means 'In all temporal worlds,  $P$ ', and 'Sometimes  $P$ ' means 'In at least some temporal world,  $P$ '. (Here we need no restriction on the accessibility relation.) The actual world amounts to the present, or Now.

Since all temporal worlds can be divided into the past, the future, and the present as the borderline, in the real *temporal (or tense) logic* people use two pairs of modal operators:

- $H(P)$ : it has always been the case that  $P$ ;
- $P(P)$ : at some time in the past,  $P$ ;

and

- $G(P)$ : it is going to be always the case that  $P$ ;
- $F(P)$ : at some time in the future,  $P$ .

$H$  and  $P$ , and  $G$  and  $F$ , are mutually definable just like  $\Box$  and  $\Diamond$ .<sup>14</sup> For the former pair, each temporal world has access only to the earlier worlds, whereas for the latter pair, each world has access only to the later worlds.

Our ordinary concept of time supports the following four biconditionals: for any  $P$ ,

- $H(P) \leftrightarrow HH(P)$ .
- $P(P) \leftrightarrow PP(P)$ .
- $G(P) \leftrightarrow GG(P)$ .
- $F(P) \leftrightarrow FF(P)$ .

For instance, the second biconditional states that  $P$  was the case at some time in the past iff there was a time point in the past at which  $P$  had been the case at some time before then. This is based on our idea that temporal worlds are *dense*,<sup>15</sup> i.e., for any two temporal worlds (or time points), there is another temporal world (time point) between them.

As you can see here, temporal logic is different from any of the three modal logics we have considered, i.e., metaphysically modal, epistemic, or deontic logic.

In the last four sections, we have considered applications of possible worlds semantics. By introducing various structures and relations (closeness, accessibility, order, denseness, etc.) among possible worlds, we can treat various different modal concepts. As I've said, there are in fact numerous modal logics out there, dealing with various modal concepts. Computer scientists, mathematicians, and linguists have been discovering more and more uses of modal logic in a broad sense. Modal logic and possible worlds semantics have become essential tools not only in philosophy but also in those areas.

### Exercise Questions

1. Explain the following four distinctions.
  - (a) Logical/non-logical truths;
  - (b) Analytic/synthetic truths;
  - (c) A priori/a posteriori knowledge and truths;
  - (d) Necessary/possible/contingent truths.
2. Explain the following concepts and distinctions.  
 Metaphysical/epistemic/deontic/temporal modality; essential/accidental properties; essence/haecceity; internal/external relations; necessary/contingent identity; de re/de dicto modality; material/indicative/subjunctive (or counterfactual) conditional; possible worlds analysis of counterfactuals; constant conjunction (or regularity) theory/counterfactual theory of causation; KK Principle.
- \*3. In Chapter 1, Exercise Question 4, which of the sentences are logical truths, and which are logical falsities?
4. If the argument from the single premise  $P$  to the conclusion  $Q$  is valid, then the conditional  $P \rightarrow Q$  is a logical truth. Also, if complex sentences  $P$  and  $Q$  are logically equivalent, then the biconditional  $P \leftrightarrow Q$  is a logical truth. Why? Go back to the definition of validity in Section 1.3 and explain.
- \*5. Consider the argument, ' $P$ . Therefore,  $Q$ '. If  $Q$  is a logical truth, the argument is valid no matter what  $P$  is. If  $P$  is a logical falsity (or

- contradiction), the argument is valid no matter what  $Q$  is. Why? Go back to the definition of validity in Section 1.3 and explain.
- \*6. For each of the following sentences, assuming that it is true, decide whether it is a logical or non-logical truth, and whether it is an analytic or a synthetic truth.
- The sky is blue or not blue.
  - All unmarried adult males are unmarried.
  - All bachelors are unmarried.
  - 21 is the legal minimum age for purchasing alcohol in the United States.
  - If 21 is the legal minimum age for purchasing alcohol in the United States, then 22-year-olds can legally purchase alcohol in the United States.
  - If 21 is the legal minimum age for purchasing alcohol in the United States, then there is a legal minimum age for purchasing alcohol in the United States.
  - If anyone at or over age 21 can legally purchase alcohol and you are 21 years old, then you can legally purchase alcohol.
  - If I am happy only if I am rich, then if I am not rich, I am not happy.
  - All happy people are happy.
  - Healthy people are happy.
7. The 45th President of the United States is Donald Trump. Is it possible for the 45th President of the United States not to be a President of the United States? Answer this by giving the *de re* and the *de dicto* interpretation of the claim.
8. What is a non-Euclidean geometry? How did the discovery of non-Euclidean geometries affect Kant's claim that there are synthetic *a priori* truths?
- \*9. Translate the following sentences in temporal logic into English, and determine, according to the ordinary conception of time, whether they are logically true (i.e., whether they are true regardless of what  $P$  and  $Q$  are). You may assume that for any  $P$ , either  $P$  or  $\neg P$  at any point of time.
- $H(P \rightarrow Q) \rightarrow (H(P) \rightarrow H(Q))$
  - $G(P \rightarrow Q) \rightarrow (F(P) \rightarrow F(Q))$
  - $\neg P(P) \leftrightarrow H(\neg P)$
  - $\neg G(P) \leftrightarrow F(\neg P)$
  - $(P(P) \wedge P(Q)) \rightarrow P(P \wedge Q)$
  - $(F(P) \vee F(Q)) \rightarrow F(P \vee Q)$
  - $P \rightarrow GP(P)$
  - $HF(P) \rightarrow (P \vee F(P))$
  - $G(P) \rightarrow F(P)$

### Discussion Questions

1. In his influential paper “Two Dogmas of Empiricism” (1951), W. V. O. Quine argued that there is no clear-cut distinction between analytic and synthetic truths, i.e., sentences true purely by virtue of meaning and sentences true at least partly by virtue of facts in the world.<sup>16</sup> Choose one common noun (such as ‘chair’, ‘dog’, or ‘human’). Can you state any clearly analytic truths involving the noun? Can you state any clearly synthetic truths involving the noun? And can you state true sentences involving the noun which are neither clearly analytic nor clearly synthetic? Is Quine correct about the blurry distinction between the analytic and the synthetic?
2. What are essential properties of human beings? What is their essence? What are essential properties of you? What is your haecceity? Discuss. What do you think about the claim that *being bipeds* is an essential property of human beings because the human DNA, i.e., the DNA of humans *as species*, enables them to stand and walk on two feet even though there are humans who cannot do so for various reasons?
3. Are Goliath and Lump1 one and the same object or two distinct objects? Discuss.
4. In Section 2.8 I introduced Kripke’s concept of rigid designator and said that since ‘Hesperus’ and ‘Phosphorus’ are rigid designators that denote one and the same object (the planet Venus) in the actual world, they must denote one and the same object in all possible worlds in which the object exists. But in Section 3.8 I introduced the distinction between necessary and contingent identity. Is it possible that ‘Hesperus’ and ‘Phosphorus’ denote two distinct objects which are only contingently identical in the actual world and distinct in other worlds? Is it plausible? Why or why not? What’s the implication of the answer?
5. Test the Trinity Thesis. Try to come up with a sentence that seems analytic but seems to express an a posteriori truth. Try to come up with a sentence that seems synthetic but seems to express an a priori truth. Try to come up with a sentence that seems to express a necessary truth that can be known only a posteriori. Try to come up with a sentence that seems to express a contingent truth that can be known a priori. How about sentences such as ‘I think’ and ‘I exist’?
6. Suppose I define ‘a primatteen’ as ‘a prime number larger than 12 but smaller than 20’. Then the sentence ‘Every primatteen is larger than 12’ seems to express the necessary truth I know a priori. Then consider the true proposition:

- (a) that the sentence ‘Every primateen is larger than 12’ expresses the necessary truth.

Is proposition (a) itself the necessary truth or a contingent truth? Do I know proposition (a) a priori or a posteriori?

Does this example refute the Trinity Thesis? Why or why not?

7. I mentioned in note 8 that Kripke also gave examples of the contingent a priori. They included cases involving the standard meter rod in Paris, the planet Neptune, and Jack the Ripper. (If you are interested, search for those in *Naming and Necessity*.) Evans (1979), however, later gave an example of the contingent a priori which is easier to deal with. In his example, we are talking about the inventor of the zip(per), but we get tired of saying ‘the inventor of the zip (if s/he exists)’, so decide to name him/her ‘Julius’ and declare:

- (b) ‘If anyone uniquely invented the zip, Julius invented the zip’.

Evans argues, following Kripke, that (b) expresses a proposition that is a contingent a priori truth. It is contingent because Julius (who is in fact Whitcomb Judson, 1846–1909) might not have invented the zip. It is a priori because we come to learn the proposition (b) expresses not by empirical investigations. Now, are you convinced that (b) in fact expresses a contingent a priori truth, thus refuting the Trinity Thesis? Why or why not?

## Suggested Further Reading

For modal logic,

- Graham Priest, *An Introduction to Non-Classical Logic*, 2nd edn.
- Lou Goble (ed.), *The Blackwell Guide to Philosophical Logic*.

The former is the best and most accessible textbook for modal logic and such non-classical logics as conditional logic and many-valued logic. The latter, though difficult, offers an extensive survey of various logics, such as deontic, epistemic, and temporal logic, and their philosophical implications.

- Paul Boghossian, “Analyticity”
- Bruce Russell, “*A Priori* Justification and Knowledge”

are careful investigations of the title topics.

- Alan Musgrave, *Common Sense, Science and Skepticism*

is an entertaining historical examination of the development of epistemology, the rationalism–empiricism debate, and skepticism in the modern period. It includes a more extensive discussion about Kant’s synthetic a priori and non-Euclidian geometry than this book, and it is also pertinent to many other issues discussed in Chapters 3 to 5.

- Robert Stalnaker, “Propositions”

gives not only a convincing criticism of Kripke’s necessary a posteriori but also a clear presentation of the theory that identifies a proposition as a function from possible worlds to truth values. Even though the paper is much neglected, it makes an ideal reading for undergraduate students.

- Michael Loux (ed.), *The Possible and the Actual*.
- Michael Rea (ed.), *Material Constitution*.
- Ernest Sosa and Michael Tooley (eds.), *Causation*.

These are useful anthologies on the title topics.

- Daniel Nolan, “Impossible Worlds”

is a compact survey of the theories of impossible worlds.

## Notes

- 1 I should note that the nature of mathematical truths is still a subject of much debate; there are probably many philosophers of mathematics who object to my characterization of mathematical truths as a species of analytic truths. Interested readers should take up philosophy of mathematics.
- 2 Some textbooks use  $\Diamond$  (tilted box) instead of  $\Diamond$  for possibility. If you forget which symbol is for necessity and which is for possibility, just remember that the tilted box  $\Diamond$  *can* fall down to the right or to the left (possibility), whereas the untilted box  $\Box$  *must* stay that way (necessity).
- 3 Following the convention in philosophy, throughout this book, when we talk about water, we will be including ice and steam.
- 4 Broadbent (2016) says that this thesis is sometimes referred to as ‘Hume’s Wall’, indicating that it is a principle Hume and other empiricists (would have) accepted. I myself have never heard of this name, however; so I stick to my naming.
- 5 Kant clearly believed this proposition even though he never used this particular example. The examples Kant in fact used include the proposition that a straight line is the shortest path between two points.



- 6 Einstein's General Relativity Theory incorporates elliptic geometry, but it is more complicated than that. According to General Relativity Theory, the shape of spacetime is not independent of the objects in it; large objects such as stars and planets bend light's path, which is considered to be a straight line, and thus distort the shape of spacetime just like grains of sand in a block of glass. The geometry of spacetime in General Relativity Theory is very complicated as a result.
- 7 At the same time, however, Locke may also have been correct when he said that we were all born with a clean slate ('tabula rasa') in the mind, if the necessary truth is the only proposition we believed when we were born. Ever since we were born, we are constantly trying to figure out what the actual world is like. When we believe proposition  $P$ , we exclude the worlds in which  $P$  is not the case (or 'the non- $P$ -worlds') from the candidates for the actual world. The more beliefs we obtain, the more we narrow down the conditions of the actual world. Since the belief of the necessary truth does not exclude any possible world in this fashion, we may consider the initial state of our mind as having only that belief. I will say more about the relation between propositions and minds in Chapter 4.
- 8 It should be noted that Kripke (1980, pp. 20–21) expressed reservations about the concept of proposition employed in the above discussion. However, he never developed a more appropriate concept of proposition (or whatever concept one can use in lieu of the concept of proposition) and used it to fill in the holes in his original argument for the necessary a posteriori. So I believe this assessment is fair, although I suspect many Kripke fans will be gravely offended by it.  
Kripke also contended that there were also contingent a priori truths. But much fewer philosophers were impressed with his argument for the contingent a priori than that for the necessary a posteriori; so we will not consider his argument for the former. I will say a little about the contingent a priori in Discussion Question 7.
- 9 Other symbolizations are  $P \Box \rightarrow Q$  and  $P \rightarrow \Box Q$ .
- 10 What I've said is (sufficiently but) not entirely accurate. More accurately, metaphysically possible worlds also have a limited accessibility relation – the equivalence relation (Section 1.5). That is, each metaphysically possible world has access only to the possible worlds in the same equivalence class. But if you are dealing only with metaphysically possible worlds, you can ignore this limitation and take the universe of possible worlds as consisting of only one equivalence class. This is what philosophers are doing (usually unconsciously) when they are talking about metaphysically possible worlds without mentioning accessibility relations. However, you mustn't ignore the limitation if you are trying to deal with different kinds of modality in one, multimodal, system.
- 11 Answer: Take any world  $u$ , and suppose  $K(P)$  holds in  $u$ . Then, in any world  $v$  such that  $R(u, v)$ ,  $P$  holds. But any such  $v$  has access to a non- $P$ -world  $w$  (i.e.,  $R(v, w)$ ) only if  $R(u, w)$  does not hold; for if it did, then  $K(P)$  would not hold in  $u$ , contrary to the initial assumption. So, on the one hand, if transitivity holds, then every  $v$  such that  $R(u, v)$  has access only to  $P$ -worlds; that is,  $K(P)$  holds in  $v$ . So  $KK(P)$  holds in  $u$ . On the other hand, if transitivity does not hold, then the

existence of a non- $P$ -world  $w$  such that  $R(v, w)$  cannot be denied by logic alone. So  $K(P)$  does not generally hold in  $v$ ; so  $KK(P)$  does not generally hold in  $u$ . Therefore, generally,  $K(P) \rightarrow KK(P)$  holds in  $u$  iff transitivity holds. QED.

- 12 Other philosophers have introduced what is virtually non-maximal possible worlds, although they usually don't call them 'possible worlds'; instead they call them *situations* (Barwise and Perry 1983). Note that once the maximality condition is dropped, only a very few propositions can have determinate truth values in a world; only a very few propositions can make up a world in that sense. For instance, there can be a world in which only the proposition that it is hot can be true and only its negation can be false; all the other propositions are indeterminate in truth value. This world may be identified as a situation that it is hot. Just like inconsistent (or impossible) worlds, whether non-maximal worlds (or situations) should exist or not is a matter of debate.

- 13 When we talked about the Ship of Theseus in Section 1.5, I said that the question whether the Original Ship of Theseus is the same ship as the Repaired/ Reassembled Ship is the question whether the former and the latter are parts of one and the same four-dimensional ship. Back then I was in fact talking about temporal parts.

The view that four-dimensional objects exist by having temporal parts is called *perdurantism*. The rival theory, *endurantism* (and three-dimensionalism) holds that each physical object exists by being wholly present (i.e., no temporal parts) at each moment of its existence. We will not discuss *endurantism* in this book.

- 14 For this reason, I love Priest's (2008) notation in which we write  $[P]$  and  $\langle P \rangle$  instead of  $H$  and  $P$ , and  $[F]$  and  $\langle F \rangle$  instead of  $G$  and  $F$ .
- 15 In fact, on our standard conception of time, temporal worlds line up not only *densely* but *continuously*. I will not bother to explain the difference, but, simply put, it is analogous to the difference between the rational number series and the real number series.
- 16 Actually, the claim Quine made is much stronger; he said that every true sentence is true partly by virtue of meaning and partly by virtue of facts – that is, there is no true sentence that is true wholly by virtue of meaning or wholly by virtue of facts. This stronger claim, however, struck many people as obviously too strong, given the existence of apparent counterexamples such as 'If it rains, then it rains', 'Every unmarried male is unmarried', and 'Every actress is female (or an actor)', which seem true purely by virtue of the meanings of the expressions involved.

# Content, Linguistic and Mental

---

### 4.1 Form and Content; Linguistic and Mental Content

So far, we have talked a lot about language, the world, and their relations. We will continue doing so throughout this book. However, we will add another element to our discussion in the next two chapters: the mind. We have already done that a little when we talked about (a priori vs a posteriori) knowledge, for knowledge is in our mind. But our discussion hereafter will be more general and include other mental states than knowledge, such as beliefs, thoughts, and desires. We will compare language with the mind, on the one hand, and investigate the relations between the mind and the world, on the other.

Every language, whether it is natural or artificial, has two facets: *form* and *content*. A system or the study of linguistic form is called *syntax* (or *grammar*), and a system or the study of linguistic content (or meaning) is called *semantics*. For instance, that neither  $P \supset Q$  nor ‘To I school go’ is a well-formed formula in their respective languages is a matter of syntax or grammar, while that ‘the most influential philosopher in antiquity’ denotes Aristotle or that ‘bachelor’ and ‘unmarried adult male’ are synonymous is a matter of semantics.

We have already discussed some important semantic concepts such as (various kinds of) extension, intension, analytic and synthetic truth, and rigid designators, but perhaps the single most important semantic concept is (simple) *truth* (as well as its opposite, *falsity*): a sentence can be true or false. *The correspondence theory of truth* holds that a sentence is true iff it corresponds to reality.<sup>1</sup> The sentence ‘Mt Everest is the highest mountain in the world’ is true iff Mt Everest is in fact the highest mountain in the world. ‘Plato was the most influential philosopher in antiquity’ is true iff Plato was in fact the most influential philosopher in antiquity. ‘Adam is a bachelor’ is true iff Adam is an unmarried adult male.

It is a curious fact that some things, i.e., linguistic expressions, obtain semantic features while other things, such as tables and chairs, do not; tables

and chairs do not denote anything and cannot be true or false. This calls for an explanation. We will consider a couple of attempted explanations in Chapter 5. One may say that a sentence expresses a proposition and that a proposition can be true or false depending on whether there is a state of affairs in the world it corresponds to. But then the question arises why some things can express propositions while others don't; why some things have (*semantic*) content while others don't.

We will not, however, pursue this issue here, and instead will focus on another interesting fact: there is another, quite different, kind of thing than linguistic expressions that also have semantic content and can be true or false. What are they? They are *mental states*. Some mental states, such as beliefs and thoughts,<sup>2</sup> have content. If I believe that Mt Everest is the highest mountain in the world, then my belief has the content that Mt Everest is the highest mountain in the world; consequently, it can be true or false, depending on whether Mt Everest is in fact the highest mountain in the world. If you think that Plato was the most influential philosopher in antiquity, then your thought has the content that Plato was the most influential philosopher in antiquity; consequently, it can be true or false, depending on whether Plato was in fact the most influential philosopher in antiquity. The content of mental states is called *mental content* while the content of linguistic expressions is called *linguistic content*. (Linguistic and mental) content is also called (linguistic and mental) *representations*. 'Representations' basically means 'copies'; what is presented in the world is represented (i.e., presented again) in language or the mind. Just like copies, pictures, and maps, representations can be correct or incorrect, accurate or inaccurate. In what follows I will use the expressions 'having content' (= 'having a representation') and 'being representational' interchangeably. So a mental state's having content is the same thing as its being representational. I may even use the redundant expression 'having representational content' to emphasize the representational nature of content.

It is a remarkable fact that such disparate things as language and mental states have content and can be true or false. It seems unbelievable that those two things are totally unrelated and obtain content independently of each other. It is more reasonable to think that one kind of content is primitive (or primary) and the other kind is derivative of (or secondary to) that kind. But which is which?

The majority of philosophers think that mental content is primitive and that linguistic content arises from mental content. One fact that lends support to this view is that some living creatures do not speak language but still seem to have mental content. Infants and dogs seem to know that the dinner is ready or that their favorite toy is in the box. Even in the case of adult humans, we do not always speak in our heads when we think. We humans seem to have developed language to convey our thoughts and desires to others and make our thoughts clearer and more precise. Language

was devised to express the content our minds have independently of, and prior to, language, these philosophers maintain.

There are, however, a minority of philosophers who believe in the primacy of linguistic content. For instance, Jerry Fodor (1975) advanced *the language of thought hypothesis*, according to which humans are born with a primitive innate language, the language of thought (or so-called ‘mentalese’), which is different from the natural languages we will learn later, such as English and Japanese. Even infants can think only by running mentalese in their heads. According to this view, mental content is nothing but representational content of mentalese.

The language of thought hypothesis is a scientific theory that can be confirmed or disconfirmed just like other scientific theories. So far, there is more evidence against it than for it. However, the final verdict has yet to be given on this issue.

## 4.2 Propositional Attitudes

I said that *some* mental states are representational and have content; I did not say *all* mental states do. Then the natural question is: what mental states do have content and what mental states do not?

The easiest (though not perfect) way to distinguish representational from non-representational mental states is grammatical. Consider:

- I believe that Mt Everest is the highest mountain in the world.
- You think that Plato was the most influential philosopher in antiquity.

These sentences have the common structure:

- Subject + mental state verb + that- (or if-)clause (sentence).

If a sentence has this form, or if it can be molded into this form without much violence to the grammar, then the mental state that the verb expresses may be considered representational; otherwise, it may not be.

For instance:

- I *believe* that Mt Everest is the highest mountain in the world.
- You *think* that Plato was the most influential philosopher in antiquity.
- We all *know* that Hesperus is identical with Phosphorus.
- I *doubt* that he will come back.
- I *wonder* if he will come back.
- Adam *sees/hears* the car running
- = Adam *sees/hears* that the car is running.
- Betty *intends* to take philosophy courses

- = Betty intends to make it to be the case that Betty takes philosophy courses.
- Charlie *wants/desires* to see a doctor
- = Charlie wants/desires it to be the case that Charlie sees a doctor.
- Charlie *wants/desires* Diane to see a doctor
- = Charlie wants/desires it to be the case that Diane sees a doctor.
- Diane wants a bagel
- = Diane wants it to be the case that Diane has/eats a bagel.
- Eli *hopes* that there will never be World War III.
- I *wish* I were a bird.
- Ponce de León *searched for* the Fountain of Youth
- = Ponce de León traveled to satisfy his desire that he find the Fountain of Youth.
- I *am happy* that I passed the exam.

All the italicized verbs above stand for representational mental states, and the that- (or if-)clauses that follow the verbs express mental content. The representational mental states are called *propositional attitudes*. That- (or if-) clauses express propositions. The representational mental states are thus attitudes toward those propositions. Hence the name.

To express the fact that the mind includes propositional attitudes, we say *that the mind has intentionality*. Reliable sources will tell you that the word ‘intentionality’ came from the Latin word ‘intentio’ and is different from ‘intention’ as we use the word today; but, certainly, the intention to do something is a typical propositional attitude. Note the difference between this ‘intention’ and the ‘intension’ I have introduced earlier as the opposite of extension. In speech, the former is called ‘*t*-intention’ (or ‘intention-with-a-*t*’) and the latter ‘*s*-intension’ (or ‘intension-with-an-*s*’). As you will see, this gets even more complicated later (sorry!) because intention and intension are connected in some interesting way. Propositional attitudes have content; that is, they are *about* things in the world, or *directed toward* things in the world. Because of that, intentionality is also called *aboutness* or *directedness*. What we call *representational mental states* may also be called *intentional mental states*.

As you may be able to see in the above list, beliefs, thoughts, and their kin make up one group of propositional attitudes, and desires, wants, and their kin make up another group. In the former case, depending on whether their content corresponds to reality, they are true or false. But in the latter case, if their content corresponds to reality, for instance, if Charlie in fact sees a doctor, we don’t say that his desire is true; instead, we say that his desire is *satisfied*. Generally, we say that content is satisfied or unsatisfied. Truth and falsity are special forms of satisfaction and unsatisfaction. Some philosophers (such as Searle 1983) think that there are (at least) two directions of fit, mind-to-world and world-to-mind, and that beliefs and thoughts exemplify

the former while desires and wants exemplify the latter. We try to make our beliefs and thoughts fit the world, whereas we try to make the world fit our desires and wants.

Back in Section 1.7, I introduced the distinction between direct and indirect quotation (remember?). There is an apparent close connection between indirect quotations and propositional attitude reports. Compare:

- Adam told me that he would go to school that day;  
Adam thought that he would go to school that day.
- Betty asked me if I had gone to school the day before;  
Betty wondered if I had gone to school the day before.
- Charlie ordered me to go to school immediately;  
Charlie wanted me to go to school immediately.

While an indirect quotation sets forth in the *that*-clause the proposition the subject expresses, a propositional attitude report sets forth in the *that*-clause the proposition the subject entertains in his/her mind. This supports the aforementioned apparent close connection between language and the mind. Again, this may be because language expresses what one has in mind (the primacy of the mind) or because propositional attitudes are a form of inner speech (the primacy of language).

Since Chapter 2, we have been identifying a proposition as a function from possible worlds to truth values. How can we think about propositional attitudes from this viewpoint? Let's call a world in which proposition *P* is true a '*P*-world'. Then when we have a belief that *P*, we are believing that the actual world is one of the *P*-worlds, excluding the non-*P*-worlds from the candidates for the actual world. The more beliefs we obtain, the more worlds we exclude, narrowing down the possibility of the actual world. Analogously, when we desire (it to be the case) that *P*, we are desiring that the actual world be one of the *P*-worlds.

Are there mental states that are not representational? If there are, what are they? Here are some potential examples:

- I have a headache.
- I am hungry (thirsty, sleepy, hot, cold, nauseated, etc.).
- I feel good.
- I am happy.

These mental states do not seem to have content. They just happen; there is no true or false, satisfied or unsatisfied here, according to some philosophers. However, the German philosopher and psychologist Franz Brentano, who reintroduced the concept of intentionality to the modern world in the late nineteenth century, maintained that all mental states

are representational, and that intentionality is the hallmark of the mental. The debate about whether there are non-representational mental states continues.

Is happiness, for instance, a propositional attitude? I included happiness in both lists of representational and non-representational mental states above. Can we just be happy without being happy about *something* (like my passing the exam)? That's a good question, but I will leave it to the reader to answer.

Some mental states have *phenomenal qualities* (often called *qualia*). Sense perceptions, for instance, carry visual or auditory images, tastes, smells, etc., which mere beliefs and thoughts do not carry. A headache also carries a phenomenal quality, a pain. Sense perceptions, however, are typical propositional attitudes while headaches are not. So the existence of phenomenal qualities seems to cut across the representational/non-representational distinction in mental states, assuming that there indeed is such a distinction. The role of phenomenal qualities in our production of knowledge is a much-debated topic in philosophy of mind (Nagel 1974, Jackson 1982, Churchland 1985).

John Locke's theory of perception is a *representational theory of perception*. According to Locke (1690), we receive ideas by perception. Ideas are copies of reality, correct or incorrect, accurate or inaccurate. Ideas of *primary qualities* (size, shape, motion, etc.) are mostly accurate copies of the qualities, whereas ideas of *secondary qualities* (colors, sounds, smells, etc.) are inaccurate, dissimilar copies of the qualities. Generally, ideas, much discussed in modern philosophy, are cousins (or grandparents?) of mental content in contemporary philosophy in many respects.

The view opposing the representational theory is called, somewhat misleadingly, *naïve (or direct) realism* (Reid 1785). The name is misleading because the theory is not a metaphysical theory about what objects exist in the world but a theory about the relation between the mind and the world, in particular, our perception of the world. According to naïve realism, we perceive the world directly as it is, and not indirectly by way of representations. As you can easily imagine, naïve realism has trouble explaining our perceptual errors and the perception's dependence on the subject's physical conditions. I see a ghost in the dark, which turns out to be just a tree in daylight. I see the river shallower than it actually is. A cold hand and a hot hand feel the same lukewarm water differently. Foods taste different when you are well and when you are sick. You take a drug, start hallucinating, and hear voices and see things that don't exist, etc. The representational theory can explain these away easily as the discrepancy between representations and reality; but naïve realism cannot. Because of that, the representational theory is currently a much more popular theory of perception than naïve realism.



### 4.3 Extensional and Intensional Contexts

That-clauses of propositional attitude reports are known to have peculiar features. In the case of a usual sentence, i.e., a sentence not involving propositional attitudes, we can substitute co-extensional expressions in the sentence without changing the sentence's truth value (i.e., the sentence's extension), as we saw earlier in Section 2.4. (The phrase 'substitutable without changing truth value' is often latinized as 'substitutable *salva veritate*', where 'verita' means 'truth' in Latin.) This conforms to the Compositionality Principle for Extension, which states that the sentence's extension is determined solely by the extensions of its constituents. If there is no change in the extensions of the sentence's parts, there will be no change in the extension of the whole sentence, either. For instance, consider:

- (1) (a) The teacher of Alexander the Great wrote *Metaphysics*;  
      (b) The most influential philosopher in antiquity wrote *Metaphysics*.
- (2) (a) Dogs are cordates;  
      (b) Dogs are renates.
- (3) The Fountain of Youth is a hot spring.

If we substitute 'the most influential philosopher in antiquity' for 'the teacher of Alexander the Great' in the true sentence (1a), the resulting sentence (1b) should remain true (even though the intension of the sentence, i.e., the proposition it expresses, may change). Both sentences are true by virtue of the fact that Aristotle wrote *Metaphysics*. Similarly, (2a) is true iff (2b) is true.

By the same principle, if a part of a sentence lacks an extension, the whole sentence will lack its extension, i.e., truth value. For instance, (3) is neither true nor false because there is no such thing as the Fountain of Youth, and, thus, 'the Fountain of Youth' in (3) has no extension. ((3) can be true/false *in the legend*, but not *in reality*.)

Neither of the above two principles generally holds in propositional attitude reports, however.

- (4) (a) Adam believes that the teacher of Alexander the Great wrote *Metaphysics*;  
      (b) Adam believes that the most influential philosopher in antiquity wrote *Metaphysics*.
- (5) (a) Betty thinks that dogs are cordates;  
      (b) Betty thinks that dogs are renates.
- (6) Ponce de León searched for the Fountain of Youth.

If Adam believes that Aristotle was the teacher of Alexander the Great and wrote *Metaphysics*, but that Plato was the most influential philosopher in

antiquity, then (4a) is true but (4b) is false. I may believe that Hesperus is the planet Venus but not believe that Phosphorus is the planet Venus, and I may believe that Mark Twain wrote *Adventures of Huckleberry Finn* but not believe that Samuel Clemens wrote *Adventures of Huckleberry Finn*; these can happen if I don't know the identities between Hesperus and Phosphorus and between Mark Twain and Samuel Clemens. Similarly, (5a) may be true but (5b) may be false. (6) is true even though 'the Fountain of Youth' has no extension.

If I kick the teacher of Alexander the Great, I am thereby kicking the most influential philosopher in antiquity because they are one and the same man. I cannot kick the teacher of Alexander the Great without kicking the most influential philosopher in antiquity. I cannot bathe in the Fountain of Youth unless the Fountain of Youth exists. Propositional attitudes are not constrained by such usual logical constraints. The mind connects to physical objects in a way different than physical objects connect to physical objects. Brentano and his followers took this as evidence that the mental cannot be reducible to the physical.

From the linguistic point of view, that-clauses of propositional attitude reports are said to create *intensional contexts* ('intension' with an 's'). A *context* is a part of a sentence, including the sentence itself. (The word 'context' can be used in different ways in other contexts – no pun intended.) Then:

- An *extensional context* is a context in which co-extensional expressions are substitutable without (even possibly) changing the extension of the whole context, and in which extensionless constituents will result in the whole context's being extensionless.
- An *intensional* (with an 's') *context* is a context which is not an extensional context, i.e., a context in which (usually) co-extensional expressions are not substitutable without possibly changing the extension of the whole context, and in which (usually) extensionless constituents may not result in the whole context's being extensionless.

If the relevant context is a sentence, then that context is extensional if co-extensional expressions are substitutable without changing the truth value of the sentence (or substitutable *salva veritate*), for the truth value is the extension of the sentence. A (usually) extensionless expression in an intensional context may not result in the absence of extension for the entire context, either. That-clauses in propositional attitude reports, then, create intensional contexts.

But why? Why do that-clauses in propositional attitude reports create intensional contexts? Frege (1892) has a neat analysis of propositional attitude reports. In a that-clause of a propositional attitude report, we do not try to describe the objective, common world; instead we try to describe the subject's conceptual world. In Adam's conceptual world, the teacher of Alexander the Great is one man (Aristotle), and the most influential

philosopher in antiquity is another man (Plato). In Betty's conceptual world, ' $x$  is a cordate' has one extension and ' $x$  is a renate' has another extension. In Ponce de León's conceptual world, the Fountain of Youth exists.

Consequently, the expressions in those that-clauses acquire uncusomary extensions (or references); specifically, what is usually the expressions' intensions (senses), i.e., what their extensions are conceived as, becomes their extensions in those that-clauses. So the extension of 'the teacher of Alexander the Great' in (4a) is not Aristotle but the individual concept Adam associates with that expression, and the extension of 'the most influential philosopher in antiquity' is another individual concept Adam associates with that expression. Thus, those expressions are not substitutable in the that-clauses in (4). Similarly, since Betty associates different properties to ' $x$  is a cordate' and ' $x$  is a renate', their extensions are different in the that-clauses in (5). 'The Fountain of Youth' expresses a certain individual concept to Ponce de León, so it has an extension in (6). Frege called these uncusomary references 'indirect references'.

As for the whole sentences in the that-clauses, in extensional contexts they denote truth values as their extensions and connote propositions as their intensions. In contrast, in intensional contexts they denote propositions and connote something else. (Frege does not say what their intensions are in such cases.)

A few paragraphs back I defined an intensional context as a context in which *usually* co-extensional expressions are not substitutable without possibly changing the extension of the whole context. The reason why I inserted 'usually' is that for Frege, intensional contexts are not really contexts in which co-extensional expressions are not in principle substitutable. For Frege, co-extensional expressions are always substitutable, i.e., substitutable both in extensional or intensional contexts. It's just that in the intensional contexts, expressions obtain unusual, non-standard extensions. Similarly, Frege's theory does not violate the Compositionality Principle for Extension, i.e., the principle that holds that the extensions of parts determine the extension of the whole.

Neat as Frege's theory may be, it did not convince everybody. To recognize one major problem with the theory, consider:

- (7) The teacher of Alexander the Great wrote *Metaphysics*, and Adam believes that the teacher of Alexander the Great wrote *Metaphysics*.

This situation makes the content of Adam's belief true, i.e., correspond to reality. According to Frege's theory, however, the semantic values of the two tokens of 'the teacher of Alexander the Great wrote *Metaphysics*' do not match; they do not say the same thing. This seems to go against our natural intuition, called 'semantic innocence' (Davidson 1968), that mental content is true or not true depending on whether it matches reality or not.

To close this section, let me point out that there is something else we are already familiar with which creates intensional contexts: de dicto modality. As we saw in Section 3.9, ‘It is necessary that the most influential philosopher in antiquity is a philosopher’, interpreted de dicto, is true; however, if we substitute for ‘the most influential philosopher in antiquity’ the (usually) co-extensional singular term ‘the teacher of Alexander the Great’, the resulting sentence ‘It is necessary that the teacher of Alexander the Great is a philosopher’ will be false. Generally, the modal operators  $\Box$  (it is necessary that ...) and  $\Diamond$  (it is possible that ...) create intensional contexts if the subsequent that-clauses are interpreted de dicto. Or, more simply, just because ‘ $2+2=4$ ’ and ‘Mt Everest is the highest mountain in the world’ have the same truth value, ‘It is necessary that  $2+2=4$ ’ and ‘It is necessary that Mt Everest is the highest mountain in the world’ do not need to have the same truth value. Obviously, treatments analogous to those for propositional attitude reports can be devised also for de dicto modal statements, though we won’t discuss them in this book.

#### 4.4 De Re and De Dicto Mental Content

Speaking of the de dicto, Quine (1956) offers an analysis of propositional attitude reports which is different from Frege’s and makes use of the de re/de dicto distinction.

In Greek mythology, Oedipus killed King Laius and then married the King’s widowed wife, Jocasta. Unbeknownst to Oedipus, they were his own biological parents. Did Oedipus want to marry his own mother? Both affirmative and negative answers are justifiable. Yes, Oedipus wanted to marry his own mother because he wanted to marry Jocasta, who is in fact his own mother. No, Oedipus did not want to marry his own mother knowingly.

Here the de re/de dicto distinction we have introduced in Section 3.9 can be repurposed. Oedipus had a de re desire, of (about, or regarding) Jocasta (the Queen, Oedipus’s mother, Laius’s wife, or however you describe her), to marry her. Oedipus did not have a de dicto desire to make the following sentence true: Oedipus marries his own mother.

Similarly, the statement ‘Lois Lane thinks that Clark Kent can fly’ can be true or false depending on whether we understand the statement de re or de dicto. Lois Lane has a de re belief, of Clark Kent (= Superman), that he can fly. She does not have a de dicto belief that the following sentence is true: Clark Kent can fly. In a previous example, Adam believes, of Aristotle (the teacher of Alexander the Great, the most famous disciple of Plato, or the most influential philosopher in antiquity), that he wrote *Metaphysics*; Adam, however, does not believe that the sentence ‘the most influential philosopher in antiquity wrote *Metaphysics*’ is true.

On the de dicto readings of (4a) and (4b), ‘the teacher of Alexander the Great’ and ‘the most influential philosopher in antiquity’ are in effect inside

quotation marks; so they do not denote the same thing, the man Aristotle; they denote different things, the expressions ‘the teacher of Alexander the Great’ and ‘the most influential philosopher in antiquity’, respectively (Section 1.7). So they cannot be substituted, and this is consistent with the Compositionality Principle for Extension.

Quine’s theory assumes that there sometimes is the *de re/de dicto* ambiguity in propositional attitude reports. As for *de dicto* reports, Quine thinks that *sentences* are tacitly involved in their content rather than *propositions*. Because of that, Quine’s theory is called *sententialism*. (Carnap 1947 advanced a similar theory.) One natural question to arise about sententialism is whether it implies that the creatures that do not speak language can have *de dicto* mental states. Quine’s answer to this question is that they can:

This semantical reformulation is not, of course, intended to suggest that the subject of the propositional attitude speaks the language of the quotation, or any language. We may treat a mouse’s fear of a cat as his fearing true a certain English sentence. This is unnatural without being therefore wrong. It is a little like describing a prehistoric ocean current as clockwise.

(Quine 1956, p. 186)

Whether Quine is ultimately correct or not, this answer seems too brief to be entirely convincing.

Epistemic modality, intensional contexts, the *de re/de dicto* distinction ... – there are many interesting similarities and connections between modality and mental content. The reason for these connections, put vaguely, seems to be that mental content creates conceptual possible worlds, worlds similar to epistemic possible worlds discussed earlier in Section 3.15.

## 4.5 Descartes’s Argument for Dualism

Whatever the reason, propositional attitude reports do create intensional contexts, i.e., contexts in which (usually) co-extensional expressions are not in principle substitutable without changing the extension of the whole context. You can use this fact to reveal the mistake Descartes made in one of his arguments for substance dualism, i.e., the view that the mind and the body are two distinct entities.

One of the major topics in philosophy of mind is *the mind–body problem*, the question about the relation between the mind and the body. Are they one and the same thing, or are they two distinct things? *Dualism*, as the name indicates, generally holds that the world is made of two kinds of things, the physical and the mental, while *physicalism* (or *materialism*) holds that the mental is nothing but the physical. In particular, *substance dualism* holds that the mind and the body – my mind and my body (or brain), in particular – are

two distinct things (substances); *property dualism* holds that, while there may only be one kind of substance in the world, the physical substance, it has two kinds of properties, mental properties and physical properties; my mind is nothing but my brain, but my brain has both mental and physical properties.

In his writing Descartes gave (at least) two arguments for substance dualism, the distinction between the mind and the body. Both arguments rely on *Leibniz's Law* (of course, it was not so called at Descartes's time), more specifically, the Indiscernibility of Identicals. I will repeat the law, which we discussed in Section 1.5:

- Leibniz's Law (the Indiscernibility of Identicals)  
For any objects  $x$  and  $y$ ,  
if  $x$  is identical with  $y$ , then  $x$  and  $y$  have exactly the same properties.

Descartes gives the following argument, the Doubt Argument, in *Discourse on Method* (1637), Part 4, and perhaps also in the *Second Meditation* (1641):

- The Doubt Argument (1)
  - I cannot doubt that my mind exists (because when I try to doubt it, I only realize that my mind is at work).
  - I can doubt that my body exists (because my perception of the body may be the evil demon's deception).
  - Leibniz's Law (the Indiscernibility of Identicals), i.e.,  $x$  and  $y$  are distinct if they have at least one different property.
  - Therefore, my mind is distinct from my body.

To put the argument into a more explicit form to which Leibniz's Law is straightforwardly applicable, define the property *existence indubitability* as the property an object has iff its existence cannot be doubted. (Dubitable = doubtable.) Then the above argument can be reformulated thus:

- The Doubt Argument (2)
  - My mind has the property *existence indubitability* (or the property  $x$ 's *existence cannot be doubted*).
  - My body does not have the property *existence indubitability*.
  - Leibniz's Law (the Indiscernibility of Identicals), i.e.,  $x$  and  $y$  are distinct if they have at least one different property.
  - Therefore, my mind is distinct from my body.

Can you see the problem with the Doubt Argument after our discussion on propositional attitudes? Yes, the problem is that doubt is a propositional attitude and thus creates an intensional context in which even some (usually) co-extensional expressions may not be substitutable *salva veritate*. The first premise of the Doubt Argument (1) is (presumably) true; that is, 'I can doubt

that my mind exists' is (presumably) false. The second premise 'I can doubt that my body exists' is true. It does not follow from these, however, that the singular terms 'my mind' and 'my body' are not usually (i.e., in extensional contexts) co-extensional, denoting two distinct entities. Even if 'my mind' and 'my body' denote one and the same thing, 'I can doubt that my mind exists' and 'I can doubt that my body exists' may have different truth values. Certainly the fact that in my conceptual world my mind and my body are separate entities does not mean that they must in fact be separate, just as the fact that in Oedipus's conceptual world his wife and his mother are separate women does not mean that they in fact are. (Descartes said that if he could clearly conceive *a* and *b* separately, God must have made them in fact separate, but this principle is obviously false.)

Thus, the Doubt Argument (1), even though its premises may all be correct, does not support the conclusion: the argument is not valid. As for the Doubt Argument (2), the first two premises are incorrect because they both incorrectly postulate the existence of the property *existence indubitability* even though there is no such property. (So when I said above, "Define the property *existence indubitability* as the property ...," I failed to define any property.) I can doubt and not doubt the existence of one and the same thing viewed from different angles; so 'existence indubitability' is not a property an object can have independently of us. (If one were to say that it is a property *in a broad sense*, then Leibniz's Law would not be universally applicable to properties in such a broad sense.) Either way, the conclusion, mind-body dualism, is not established by the Doubt Argument.<sup>3</sup>

Contrast this with another argument for dualism Descartes gives in the *Sixth Meditation*, which I call 'the Extendedness and Divisibility Argument' (or 'E&D Argument', for short). It goes like this:

- The E&D Argument
  - My body is extended and divisible.
  - My mind is not extended or divisible.
  - Leibniz's Law (the Indiscernibility of Identicals), i.e., *x* and *y* are distinct if they have at least one different property.
  - Therefore, my body is distinct from my mind.

Unlike the Doubt Argument, this argument is a valid argument. That is, if 'My body is extended and divisible' is true but 'My mind is extended and divisible' is false, then 'my body' and 'my mind' must denote different objects. That's because those sentences create extensional contexts. As is often pointed out, the problem with the E&D Argument is the second premise; despite Descartes's assertion, it is not obvious that the premise is true. But I won't get into that here. My point, rather, is that the Doubt Argument has a feature related to the intensionality of propositional attitudes that the E&D Argument does not have. They are problematic arguments for different reasons.

## 4.6 Skepticism; 'a Brain in a Vat'

Descartes was a (methodological) skeptic.<sup>4</sup> His evil demon argument is a skeptical argument. But what is skepticism?

Generally, skepticism is an attitude of doubt. If one is skeptical about  $x$ , one is doubtful about  $x$  (its existence, etc.). So one may be skeptical about something specific, e.g., the existence of the Fountain of Youth, the Loch Ness monster, Santa Claus, the Tooth Fairy, etc. One may also be skeptical about the morality of abortion, euthanasia, war, etc. Ponce de León's belief and desire content includes the Fountain of Youth. But, as we already know, something's existing in one's mental content is no guarantee that it exists in reality. So skeptics about the Fountain of Youth doubt that Ponce de León's mental content involving the Fountain of Youth corresponds to reality. Skeptics about the morality of abortion doubt that some people's belief that abortion is a morally permissible act corresponds to reality. Religious agnostics doubt that many people's belief in God corresponds to reality.

In philosophy, however, skepticism is often more wholesale. It may cast doubt over an entire subject, such as the external world (i.e., the objective physical world existing outside one's mind), other minds, morality, religion, etc. For instance, the skepticism about the external world casts doubt on the existence of the entire external world, not just about the existence of specific objects such as the Fountain of Youth and Santa Claus. The skepticism about morality casts doubt about the objective correctness and incorrectness of moral judgment in general, not just about the morality of specific acts such as abortion or euthanasia.

Descartes and David Hume raised skepticism about the existence of the external world. In the *First Meditation* Descartes raised the possibility that a powerful and cunning evil demon exists and keeps on deceiving him into thinking that the external world exists while, in reality, it does not (though he erased this doubt later). Hume argued that Locke's representational theory of perception allows us to have access only to ideas of the external world and never to the world itself; thus, we cannot deny the possibility that the external world does not exist. Kant turned this skepticism into an even more radical thesis about the absolute unknowability of Noumena.

The skepticism about  $x$  maintains that we never know whether  $x$  is the case (or  $x$  exists) or  $x$  is not the case ( $x$  does not exist). Berkeley considered himself not a skeptic about the external world and in fact argued against the skepticism (in *Three Dialogues between Hylas and Philonous in Opposition to Sceptics and Atheists*, 1713) because he was convinced that the material world does not exist. His position, called *immaterialism*, strictly speaking, is not skepticism but simply anti-realism about the material world. Analogously, atheism is not skepticism but simply anti-realism about the existence of God.



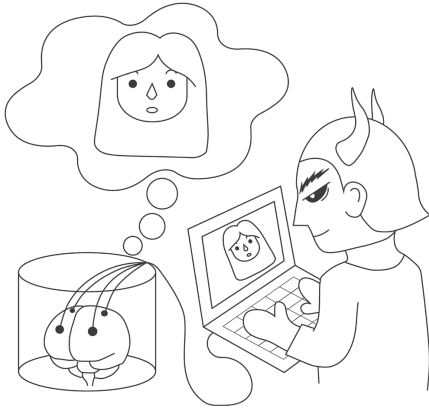


Figure 4.1 A brain in a vat.

A denial of any standard, commonsensical position, however, is often labeled skeptical in a broad sense.

The *brain in a vat hypothesis* is a contemporary version of Descartes's evil demon hypothesis. In this hypothesis (Figure 4.1), my brain is placed in a vat, kept alive, and keeps on receiving signals from a supercomputer wired to it, which is controlled by a mad scientist (a contemporary cousin of the evil demon). As a result of the signals, the brain keeps on experiencing the kinds of things I experience in normal circumstances, such as the perception of my body moving around in the physical environment, and is thinking that I have a physical body and am conducting a normal life.

The original challenge is to answer the skeptical question: how do I know that I am in fact not a brain in a vat? We will examine one answer, given by Hilary Putnam, later in Section 5.13.

This 'brain in a vat' scenario, however, has been used for various purposes in various areas of philosophy, not only in epistemology but also in philosophy of mind and philosophy of language, or even in ethics. Just like possible worlds, it is one of philosophers' favorite toys. Along with Descartes's original dream and evil demon arguments, it is also the source of numerous popular movies, such as *The Matrix*, *Inception*, *Total Recall*, and *Eternal Sunshine of the Spotless Mind*, to name a few.

Unlike Descartes's original skeptical arguments, the brain in a vat hypothesis does not deny the existence of the entire external world; some physical objects, such as the brain, the vat, the computer, and the mad scientist, are assumed to exist. However, it is still a source of wholesale skepticism because it assumes that physical reality, though it in fact exists, is radically different from what we think it is.

## 4.7 Moral Error Theory

Another area of philosophy where skepticism has a strong foothold is ethics. *Moral error theory* (Mackie 1977) is a version of moral anti-realism, the position that denies the existence of moral properties and moral facts. Moral error theory, however, does not deny that our judgments and assertions about morality have representational content. It's just that they are fundamentally mistaken; hence the name 'error theory'. If somebody says 'Abortion is wrong' and another says 'Abortion is morally permissible', both are incorrect, according to moral error theory, because there is no such property as *being morally wrong* or *being morally permissible*, to be instantiated by abortion or any other acts. More accurately, their claims are neither true nor false because parts of them lack denotation.<sup>5</sup>

There is reason to be more skeptical about the existence of moral properties than non-moral properties such as the property *being green*. There is a good reason to think that when I think that the traffic signal in front of me is green, this mental content usually corresponds to reality and the traffic signal in front of me is in fact green. The scientific explanation of the way I obtained this belief begins with the state of affairs that the traffic signal in front of me emitted light waves of wavelength 500–550 nanometers, i.e., green light (Figure 4.2). These light waves hit the retinae of my eyes, which in turn sent neural signals to my brain, which produced the mental content that the traffic signal in front of me is green. If the light waves the signal emitted had been of wavelength 650–700 nanometers instead, then I would have obtained a different belief, i.e., that the signal is red. If the signal emitted green light but I obtained the belief that the signal is red, then the explanation will be complex, but would still begin with the state of affairs that the signal emitted green light. (Something unusual happened afterwards.) Here that

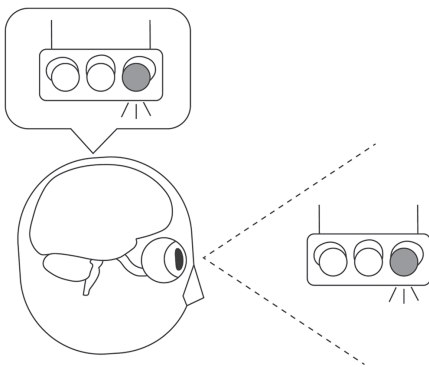


Figure 4.2 'The signal is green'.

my mental content was correct and that the signal in front of me was in fact green was a big deal, for that correlated with the fact that the signals at the right and left were red. As a result, my car safely crossed the street without being hit by another car from the side. In sum, we have reason to believe that the property *being green* exists because it enters into the explanation of why we obtain the mental content that certain things are green.

That's not the case with moral properties: moral properties do not enter into the explanation of why we obtain the mental content that certain things have those moral properties. The explanation of why Adam believes that abortion is morally wrong while Betty believes that abortion is morally permissible does not begin with the state of affairs that abortion is morally wrong or that abortion is morally permissible. The explanation is a very complex one, citing Adam's and Betty's upbringing, social environment, education, etc., but the alleged fact that abortion is morally wrong or that it is morally permissible, whichever the case may be, does not play any role in the explanation (Figure 4.3).

To use a slightly technical term, unlike the state of affairs that the signal is green, the alleged state of affairs that abortion is morally wrong/permissible does not have 'causal efficacy' (or, more simply, causal power) to the creation of the belief that abortion is morally wrong/permissible.

Make no mistake: *our belief* that abortion is morally wrong/permissible has much causal efficacy. We act on that belief, joining demonstrations, shouting out in the classroom, or bombing an abortion clinic, for instance. The error theorist's point, however, is that the alleged state of affairs that abortion is morally wrong/permissible does not have causal efficacy to the creation of that belief. We can explain those actions without assuming the existence of the state of affairs that abortion is morally wrong/permissible. In contrast, the state of affairs that the signal has just turned green has itself causal efficacy, not only creating my belief that the signal is now green, but also zapping the insects to death which were swirling around the light, for instance.

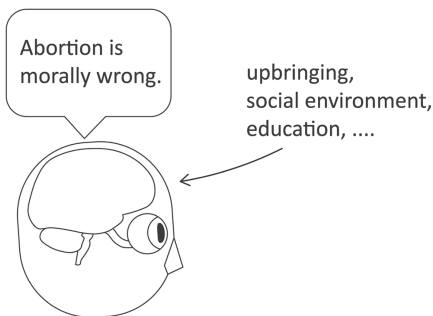


Figure 4.3 'Abortion is morally wrong'.

The following modal consideration should further bolster the claim that moral facts do not have causal efficacy. Think of three possible worlds. The original world contains numerous natural (i.e., non-moral) facts, including natural facts related to abortion to the minutest detail: who has an abortion when and where, who avoids abortion, who performs abortion to whom, who is for and who is against abortion for what reason, what kind of political system the country has, what the laws are concerning abortion, etc. There are, however, no moral facts in this world, no moral facts about abortion in particular. Now, add to this world the moral fact that abortion is morally wrong – this is the second world. Add to the original world the moral fact that abortion is morally permissible – this is the third world. However, we seem to be able to do this too easily, without changing any non-moral facts. That is, these moral facts do not affect any existing non-moral facts; they don't have causal efficacy.

In reaction to this line of argument, one may say that perhaps the state of affairs that abortion is morally wrong (assuming that it is) exists really as a natural state just like the state that the signal is green, and causally affects our moral sense, just as the state of affairs that the signal is green affects our perception; it's just that scientists have not yet discovered the causal mechanism. However, one strange thing about the alleged moral states of affairs which natural (or non-moral) states of affairs do not share is that moral states of affairs imply some 'ought' (or 'should') facts, i.e., the facts about what we morally ought or ought not to do. The alleged state of affairs that abortion is morally wrong implies the state of affairs that we ought to avoid abortion (unless there is an overriding reason we ought not to). It is widely believed that science can tell you what *is* the case, but not what (morally) *ought* to be the case. So it seems highly unlikely that the state of affairs that abortion is wrong exists really as a natural state and causally affects us.

## 4.8 Performative Utterances

Note that moral error theory sets forth two theses. The first is moral anti-realism, i.e., the metaphysical thesis that there are no (objective) moral properties and facts in reality. The second is the semantic thesis that our judgments and assertions about moral properties and facts have genuine representational content. The consequence of these two theses is that our judgments and assertions about moral properties and facts inevitably fail to match reality. This is analogous to the previous example in which Ponce de León's beliefs about the Fountain of Youth inevitably fail to match reality because there is no such thing as the Fountain of Youth. In both cases, we inevitably fail to hit the target because the target is an illusion.

Now, if moral error theory is correct and our judgments and assertions about the morality of issues like abortion always fail to match reality, the

intellectually honest thing to do seems to be to stay away from having those judgments and assertions, e.g., not to believe or say either that abortion is wrong or abortion is permissible. Not to engage in moral discourse.

Many moral anti-realists, however, think that's a mistake; a mistake based on a misunderstanding about the nature of moral discourse. Specifically, they maintain, assertions about moral properties and facts do not have representational content. They do not even aim at matching reality. This view is called *moral expressivism* (or *moral non-cognitivism*). Moral expressivism is a version of moral anti-realism, embracing the first, metaphysical, thesis above, but denying the second, semantic, thesis. I will discuss moral expressivism in the next section.

But how is such a view possible? 'Abortion is morally wrong' is a declarative sentence (Section 1.3). How can a declarative sentence not have representational content? I would like to answer this question first in this section.

The answer comes from *speech act theory*. Speech act theory was first developed by the Oxford philosopher J. L. Austin in the mid twentieth century. You may have noticed that so far in this book, I've treated natural language almost like an autonomous machine: it has its own syntax and semantics, and if an expression conforms to the grammar, it will automatically receive an intension as well as an extension (if there is one). What's missing in this picture, according to speech act theorists, is the human speaker. Put simply, speech act theory sees language not as an autonomous formal system but as a tool humans created and use for various purposes – a part of human acts. Austin's influential lecture series, later edited into a book, is appropriately titled *How to Do Things with Words* (Austin 1962).

Among other things, Austin introduced the concept of *performative utterance*. For instance, consider utterances of the following sentences:

- I apologize.
- I promise.
- I order you to do fifty push-ups.
- I do (i.e., I take this woman to be my lawfully wedded wife).
- I now pronounce you husband and wife.
- I sentence you to death.
- I name this ship 'Queen Elizabeth II'.
- I declare war.
- You are fired.

If such utterances are made by proper individuals in proper circumstances (e.g., by the priest in a church wedding ceremony or by the judge at a trial), then they will indeed become acts of apology, promise, order, marrying, sentencing, naming, etc. I am indeed apologizing, promising, or ordering when I say 'I apologize', 'I promise', or 'I order you to do fifty push-ups'.

This is suggested by the fact that I can add the adverb ‘hereby’ to all of those sentences, indicating that I am thereby performing that very act. It is a mistake to think that when I say ‘I apologize’, I am trying to describe my sorry feeling or an independently existing act of apology; I may be *expressing* my sorry feeling (in the expressivist sense) but not *describing* it. This is quite different from the third-person utterance of ‘He apologized’ or the first-person utterance of the past tense ‘I apologized’ or the present progressive ‘I am apologizing’, each of which is a genuine description of my act of apology. Even utterances like ‘Sorry’ and ‘Thank you’ are not descriptions of any independent facts but themselves acts of apology or thanking. It is a mistake to think that every declarative sentence aims at describing an independently existing fact; Austin called it ‘the descriptive fallacy’.

Austin maintained that sentences uttered in performative utterances like the above do not have representational content and thus are not truth-valuable or ‘truth-apt’. This is different from, e.g., the fact that the sentence ‘The Fountain of Youth is a hot spring’ lacks a truth value because a part of it does not have a denotation; this sentence is still truth-apt, i.e., could be true or false if the Fountain of Youth existed. There are successes and failures in performative utterances. For instance, if I am not the right person in the right circumstances, my utterance does not count as an act of marrying, sentencing, naming, or declaring war. But that’s not the same thing as making a false statement, Austin argued.

## 4.9 Moral Expressivism and the Frege-Geach Problem

Moral expressivism (or moral non-cognitivism) considers moral statements such as ‘Abortion is morally wrong’ performative utterances of sort: they do not have representational content and are not truth-apt. They do not *describe* the speaker’s opinion or cognitive mental state; instead, they *express* his/her non-cognitive attitude toward the subject. There are several versions of moral expressivism. According to *emotivism* (Ayer 1936, Stevenson 1937), moral statements are expressions of the speaker’s emotions. Thus, the statements ‘Abortion is morally wrong’ and ‘Abortion is morally permissible’ are similar to the exclamations ‘Boo, abortion!’ and ‘Hooray for abortion!’, respectively. According to *prescriptivism* (Hare 1952), moral statements are more analogous to imperatives; so ‘Abortion is morally wrong’ and ‘Abortion is morally permissible’ are similar to ‘Don’t have (or perform) an abortion’ and ‘Don’t avoid abortion’, respectively.

To many moral anti-realists, expressivism seems to be a more attractive option than error theory, for it allows them to engage in moral discourse, for instance, discussions about what is morally right and what is morally wrong, even though there are no objective moral facts in reality. Of course, there are no ultimately correct answers to such moral questions because

there is no fact of the matter. Moral discourse will have to be nothing but a sophisticated form of shouting match, one side's hooraying and the other side's booing for the subject or each side's telling the other side what to do. Still, that's better than having to be completely silent.

However, moral expressivists have had a hard time defending the claim that moral statements do not have representational content. 'Abortion is wrong' does not strike us as grammatically much different from 'The signal is green' or, perhaps even better, 'The signal is wrong (i.e., malfunctioning)'. How can the former be non-representational while the latter is representational? That just does not seem possible.

True, this itself is a weak argument appealing to one's raw linguistic intuition, but the British philosopher Peter Geach (1965) has formulated a more sophisticated, convincing argument based on Frege's work, showing that the meanings of moral sentences like 'Abortion is wrong' cannot be essentially different from the meanings of non-moral sentences like 'The signal is green' or 'The signal is wrong'. The problem with moral expressivism that his argument points out is thus called *the Frege-Geach problem*. It is also called *the embedding problem*, for it has something to do with the meanings of logically complex sentences which contain moral sentences as their constituents. It has much to do with Frege's Compositionality Principle; hence the name.

For instance, consider:

- (a) Abortion is not wrong (i.e., it is not the case that abortion is wrong).
- (b) Abortion is wrong or abortion is not wrong.
- (c) If abortion is wrong, fetuses must be humans.

If the atomic sentence 'Abortion is wrong' does not have representational content and is not truth-apt, how can we make sense of the possible truth of these complex sentences? As you can see in (a), 'Abortion is wrong' can sensibly be negated, but 'Boo, abortion!' cannot be; there is no such thing as the negation of an exclamation. The negation of an order also makes little sense. Obviously, 'Have an abortion' should not be the negation of 'Don't have an abortion', for 'Abortion is not wrong' cannot be the same as the order 'Have an abortion'. And the difficulty multiplies if we consider more complex sentences. Even if we should find an emotive or prescriptive meaning of 'Abortion is not wrong', how could we make a disjunction with the emotive or prescriptive meaning of 'Abortion is wrong'? 'Boo! or OK!' makes no sense. In contrast, (b) makes perfectly good sense and is an apparent logical truth. Similarly, (c) makes perfectly good sense, whereas neither 'Boo, abortion!' nor 'Don't have an abortion' can be used as the antecedent of a conditional.

One might try to avoid these problems by maintaining that the sentence 'Abortion is wrong' has a different, descriptive, meaning in complex sentences

than it does when it is used in isolation. This tactic, however, does not work. Consider the following argument:

- (d) – If abortion is wrong, then fetuses must be humans.
  - Abortion is wrong.
  - Therefore, fetuses must be humans.

This argument is clearly valid. However, it could not be if ‘Abortion is wrong’ in the first premise and that in the second premise had different meanings. So ‘Abortion is wrong’ must have the same meaning whether it is used in isolation or as a part of a complex sentence.<sup>6</sup>

Some people may wonder if we can simply assign ‘neither true nor false’ to moral sentences such as ‘Abortion is wrong’. The resulting logic, which allows the truth value gap, would be complicated, to be sure, but that’s no different from the case involving ‘The Fountain of Youth is a hot spring’, which is also neither true nor false.

However, there is a big difference. ‘The Fountain of Youth is a hot spring’ *can* be true (or false), i.e., there is a possible world in which it is true (or false), i.e., a possible world in which the Fountain of Youth exists and is (or is not) a hot spring. In contrast, ‘Abortion is wrong’ *cannot* be true or false, according to expressivism; it is not the kind of thing that can be true or false – it is not truth-apt – just like tables and chairs. And the following is a consequence of this difference. Compare the arguments:

- (e) – The Fountain of Youth is a hot spring.
  - Therefore, snow is red.
- (f) – Abortion is wrong (or permissible).
  - Therefore, snow is red.

Recall the definition of valid argument given in Section 1.3: An argument is valid iff it is impossible for all its premises to be true and its conclusion false. Applying this definition, (e) is *not valid*, as we expected it would not be, because it is possible for its premise to be true (i.e., the Fountain of Youth exists and it is a hot spring) but its conclusion false (i.e., snow is not red). So it is reasonable for us to believe that the Fountain of Youth is a hot spring without also believing that snow is red. In contrast, (f) ought to be *valid*, contrary to our expectation, for it is not possible for its premise to be true and its conclusion false, because it is not possible for its premise, ‘Abortion is wrong (or permissible)’, to be true, regardless of the conclusion. So it is not reasonable for us to believe that abortion is wrong/permissible without also believing that snow is red; if we believe that abortion is wrong/permissible, we also must believe that snow is red. Generally, on the proposal under consideration, any moral sentence would imply any sentence as true; ‘Abortion is wrong (or permissible)’. Therefore, *P* would be valid for any



sentence *P*. This is a crazy consequence we cannot accept.<sup>7</sup> Therefore, we cannot just brush off the Frege-Geach problem by assuming that moral sentences are neither true nor false.

In conclusion, 'Abortion is wrong' and other moral sentences must have representational content and must be truth-apt just like non-moral sentences such as 'The signal is green (or wrong)'. Because of this argument, many contemporary moral expressivists have given up the idea that moral statements do not have representational content. It is beyond the scope of this book to discuss how they still might (or might not) be able to retain the spirit of expressivism.

In closing, I should note that basically the same argument as above can be given against Austin's claim that performative utterances of sentences like 'I apologize' and 'I order you to do ...' do not have representational content and thus are not truth-apt. All of the following sentences apparently do have representational content and can be true or false:

- I don't sentence you to death.
- I don't name this ship 'Queen Elizabeth II'.
- You are not fired.
- I order you to do fifty push-ups or I don't order you to do fifty push-ups.
- If I apologize, then I will have to pay for the damages.
- Even if I now pronounce you husband and wife, you won't be until you sign legal documents.

The same argument as above will show that sentences like 'I apologize' and 'I order you to ...' should have representational content and be truth-apt.

Furthermore, 'I apologize' also has connections to 'I apologized', 'I am apologizing', and the third-person description 'S/he apologized'. Suppose that Adam says to Betty 'I apologize' and that Betty says to Charlie 'Adam apologized'. If the latter is a description of Adam's act of apology, which is his utterance of 'I apologize', then why shouldn't the former also be a description of the same act, i.e., his utterance of those very words? There is an obvious gap between the premise and the conclusion of the following argument:

- 'I apologize' does not describe my independently existing act of apology.
- Therefore, 'I apologize' does not describe my act of apology.

The following is still a possibility:

- 'I apologize' describes my act of apology, which is my utterance of that very sentence.

Thus, there is no reason to think that 'I apologize' and other sentences uttered in performative utterances do not have representational content and are not truth-apt.

Indeed, while performative utterances often make the uttered sentences true automatically, that's not always the case and there are uttered sentences that seem false. 'I name this ship "Queen Elizabeth II"' and 'You are fired' seem false if I am not in a position to name the ship or fire anybody, for in that case I in fact do not name the ship 'Queen Elizabeth II' and you are in fact not fired.

In the representational picture I have depicted in this chapter, declarative sentences and certain mental states express propositions, which will be true or false or neither depending on whether they correspond to the states of affairs existing in reality; consequently, the sentences and mental states will be true or false or neither, too. While speech act theory and the concept of performative utterances do give us a new way of looking at our use of language, it does not seem to force revision to this picture. Whether it is ultimately correct or not, the picture stands as a default position at least for now.

### Exercise Questions

1. Explain the following concepts and distinctions.  
Form/content; syntax (or grammar)/semantics; linguistic/mental content; representations; language of thought hypothesis; propositional attitudes; intentionality; phenomenal qualities (or qualia); representational theory of perception/naïve (or direct) realism; extensional/intensional contexts; de re/de dicto mental content; sentimentalism; mind/body problem; dualism/physicalism (or materialism); substance/property dualism; skepticism; moral error theory/expressivism (or non-cognitivism); emotivism/prescriptivism; speech act theory; performative utterances; Frege-Geach problem.
- \*2. State the mental content involved in the following cases. If there is none, state 'Non-representational'. Some contentious cases may be included.
  - (a) I believe that the world is flat.
  - (b) She has a headache over your yelling.
  - (c) Adam is looking for someone to love him.
  - (d) Betty is happy that you came.
  - (e) Betty is happy because you came.
  - (f) I saw the truck run over the dog.
  - (g) Charlie wants to be rich.
  - (h) I am in pain from my scraped knee.
  - (i) I wonder if you will come to the party.

- \*3. Which of the following are performative utterances? Some contentious cases may be included.
- (a) 'I love you'.
  - (b) 'I accept your apology'.
  - (c) 'The class (is) dismissed'.
  - (d) 'I quit' (in the present tense).
  - (e) 'I quit' (in the past tense).
  - (f) 'You are under arrest'.
  - (g) 'Adam is under arrest'.
  - (h) '(You are) out!' (A call made by the home-plate umpire in a baseball game.)

### Discussion Questions

1. Is happiness a propositional attitude? Do I have to be happy about *something*, like passing the exam? Or can I be just happy, not happy about anything in particular? Can there be two kinds of happiness, one about something and the other not about anything in particular? Does happiness have a distinctive phenomenal quality?
2. Related to Q.1. I can say 'I am pleased (surprised, etc.) that I passed the exam'. I can also say that 'I am pleased (surprised, etc.) *to learn (or hear, see, etc.)* that I passed the exam'. Is the former just an abbreviation of the latter? If it is, that may suggest that pleasure, surprise, etc., are not independent propositional attitudes – or not even propositional attitudes at all – but non-representational mental states that arise as a result of the propositional attitude *learning* (= coming to believe). What do you think? Can you say the same thing also about happiness?
3. As we saw in Section 4.3, one feature of a propositional attitude is that the content of the attitude may not be true and that the object the attitude is directed toward, e.g., the Fountain of Youth, may not exist. Now, people whose limbs have been amputated sometimes experience phantom pain, i.e., pain that feels like it is coming from the body part that's no longer there. Someone may feel that his leg is hurting, but the leg no longer exists. Does this show that having pain is a propositional attitude? Why or why not?
4. As regards Quine's defense of sententialism, how could something, like an animal or an infant, that does not understand the relevant language have an attitude toward sentences in that language in the way Quine describes? Do you find sententialism plausible? Which do you find to be a better solution to the substitution problem, Frege's theory or Quine's sententialism?

5. In Section 4.3, I mentioned Brentano's argument for dualism. Using the format given in Section 4.5, it can be expressed as follows:
  - The Argument from Intentionality
    - My mind is related to the material world in the way describable only in intensional contexts.
    - My body is not related to the material world in the way describable only in intensional contexts. My body's relation with the material world is describable entirely in extensional contexts.
    - Leibniz's Law (the indiscernibility of identicals).
    - Therefore, my mind is distinct from my body.

What do you think about this argument? Are you convinced by this argument that dualism is correct? Why or why not?

6. In Section 4.7, I gave an error theorist argument that there is no state of affairs that abortion is wrong/permissible that causes our belief that it is. Do you agree with the argument? State the points of agreement and disagreement. If you have people who have points of disagreement, discuss those points.

## Suggested Further Reading

One of the best introductions to philosophy of mind is

- Ian Ravenscroft, *Philosophy of Mind: A Beginner's Guide*.

On propositional attitudes,

- Michael Nelson, "Propositional Attitude Reports"

is a thorough survey.

- J. L. Mackie, *Ethics*

is a very accessible introduction to ethics which also sets forth Mackie's moral error theory.

- J. L. Austin, *How to Do Things with Words*, 2nd edn

is an enjoyable original source for Austin's speech act theory.

## Notes

- 1 Two other major theories of truth are *the coherence theory* and *the deflationary (or disquotational) theory*, but we won't discuss those theories in this book. The coherence theory of truth is different from coherentism in epistemic justification, to be discussed in Section 5.11. I will say just a little (but not much) about the deflationary theory in Section 5.13.
- 2 Here the word 'thought' is used in an informal, non-Fregean fashion, not to be identified with propositions. Recall (Section 2.4) that Frege called what we call 'propositions' 'thoughts'.
- 3 It must be noted, however, that some influential contemporary philosophers (Kripke 1980; Chalmers 1996, 2010) defend dualism along Descartes's lines.
- 4 Of course, Descartes was not a *genuine* skeptic. In the end he accepted the existence of the material world. But in the *First* and *Second Meditation*, Descartes tried to call everything into question in order to determine what cannot be called into question. This method of his is called *methodological skepticism (or skeptical method)*.
- 5 Claims such as 'Abortion is *not* wrong' are trickier. They are ambiguous. 'Abortion is not wrong' can be understood in two ways: it can be understood as 'Abortion is not-wrong (i.e., permissible)', or it can be understood as 'It is not true that abortion is wrong'. According to error theory, the former is neither true nor false because the property *not being (morally) wrong*, just like the property *being (morally) wrong* or *being (morally) permissible* does not exist; but the latter is true because 'Abortion is wrong' is certainly not true (or false). To use the terms introduced earlier (Chapter 2, note 9), the negations used in the former and the latter are an internal negation and an external negation, respectively. The difference between the present case and 'The Fountain of Youth is not a hot spring' (or 'The present King of France is not bald'), discussed there, is that while 'The Fountain of Youth is a hot spring' is neither true nor false because the singular term 'the Fountain of Youth' lacks denotation, in the present case 'Abortion is wrong' is neither true nor false because the predicate '*x* is (morally) wrong' lacks denotation (does not denote even the empty set). This is possible if some sparse theory of properties is correct. (A possible alternative view is that moral properties such as *being (morally) wrong* and *being (morally) permissible* do exist, but that nothing instantiates them; so, for instance, 'Abortion is wrong' and 'Abortion is permissible' are both false. We will not consider this variant of error theory in this book.)
- 6 You may notice some resemblance between this argument and the argument given at the end of Section 4.3 against Frege's analysis of propositional attitude reports. That argument suggests that Frege violated his own principle in his analysis of propositional attitude reports.
- 7 We may be able to avoid this problem if we change our logic to the kind of non-classical logic I mentioned in Section 3.15; however, it seems highly ad hoc to make such a radical change only for that purpose.

# Internalism and Externalism

---

### 5.1 Internalism vs Externalism in General

We make the distinction between internalism and externalism in various areas of philosophy such as metaphysics, epistemology, philosophy of language, philosophy of mind, and ethics. You may have heard about some instances of it. Even though internalism and externalism are always internalism/externalism *about* something, some specific topic, such as internalism/externalism *about epistemic justification* or *about linguistic meaning*, there is good reason why those views are characterized by the same pair of terms ‘internalism’ and ‘externalism’.

Philosophers talk about internalism vs externalism about things; but internal/external *to what*? The answer is, usually, *to the mind*.<sup>1</sup> So, for instance, internalism about epistemic justification holds that whether you are justified in believing a certain proposition or not is determined purely by what you have in mind, whereas externalism about epistemic justification denies that, maintaining that whether you are justified in believing something is determined at least partially by things outside of your mind – physical surroundings, for instance. Similarly, internalism about linguistic meaning, or *meaning internalism*, holds that the meaning of any linguistic expression you use is determined purely by what is in your mind, for instance, the concept you attach to the expression, whereas *meaning externalism* holds that meaning is determined at least partially by elements outside of the mind. Internalism about reasons, or *reasons internalism*, holds that reasons for an action must always motivate the agent to perform that action, whereas *reasons externalism* holds that there can be objective reasons for an action which are completely disconnected with the agent’s motivations to perform that action. For instance, you may have an internal reason to avoid an abortion, e.g., your desire to have a family; but there may also be an external reason to avoid an abortion, e.g., the fact that abortion is morally wrong (assuming that it in fact is). While the former reason is in your mind and affects your action, the latter reason, though it is still an objective reason, is not in your mind and does not affect your action, according to reasons externalism.

Furthermore, it is usually assumed that you have access to whatever you have in your mind: you can get it on reflection. Thus, ‘the subconscious’ – something that may be thought to exist at the bottom of your mind, something you yourself don’t have access to – is not ‘in your mind’ in this sense. The subconscious is as external to you as your physical surroundings in this respect, for you cannot use it as you wish. So internalism about epistemic justification usually implies that you yourself can know whether your beliefs are justified, whereas externalism about epistemic justification says that you may not know. Similarly, you may not have access to external reasons for your action; it may even not occur to you that abortion can be morally wrong. Another implication of externalism in general is that even if two people may have exactly the same mental states, they may have different status with regards to the topic, for instance, one may be justified but the other unjustified in believing the same thing, or they may mean different things by the same word.

In this chapter, we will consider three different instances of the internalism/externalism distinction in three parts. Part A will discuss internalism vs externalism about linguistic meaning and mental content, in particular about the meaning of proper names and natural kind terms. A few concepts related to the distinction, such as intrinsic and extrinsic properties and supervenience, will be explained toward the end of the part. The main topic of Part B is internalism vs externalism about epistemic justification. Along the way, a few very basic theories in epistemology will be introduced, such as the JTB theory of knowledge, foundationalism, coherentism, and reliabilism. Part B ends with a section about Putnam’s semantic externalist argument against the ‘brain in a vat’ skepticism. Part C concludes the book with a discussion about John Locke’s famous internalist theory of personal identity, which holds that personal identity consists in our memory or stream of consciousness. We will see this theory undermined by the need for external elements in determining personal identity.

## **PART A: SEMANTIC INTERNALISM AND EXTERNALISM**

### **5.2 The Description Theory of the Reference of Proper Names**

Language has an amazing power. If I say ‘Aristotle’ or ‘the teacher of Alexander the Great’, these word tokens (sounds or inscriptions) somehow reach back almost twenty-five centuries to the past, cross the Atlantic, and connect to their reference, the ancient Greek philosopher Aristotle. If I say ‘the Big Bang’, it reaches back more than 13 billion years and connects to the event that created the whole universe. No time machine can do

that! Then the natural question is: What is the mechanism of reference that enables tokens of singular terms, mere sounds or smidges of ink, to connect to their references?

Here merely saying that each of those tokens expresses a certain individual concept (Section 2.4), and that anything that satisfies that concept will be the token's reference, cannot be the whole answer, for it then calls for an explanation about how the token can express that particular individual concept. We need a deeper answer than that.

Recall that, setting aside demonstratives, there are two kinds of singular terms: proper names and definite descriptions. In the case of definite descriptions such as 'the teacher of Alexander the Great', the answer to the above question actually seems pretty obvious. As soon as we understand the expression 'the teacher of Alexander the Great', we understand how to single out its reference; we know what to look for if we were asked to find the reference of the term: whoever is teaching Alexander the Great. So, setting aside the practical problems of how to go back to ancient Greek and get into Alexander's palace, theoretically we can reach the term's reference. This ability of ours is what bridges the enormous spatiotemporal gap between the uttered term and its reference. Basically the same thing can be said about other definite descriptions. This is obviously an internalist account because, according to this account, what bridges a definite description and its reference is the speaker's understanding of the description, and understanding is in the speaker's mind.

However, it is much more difficult to answer the question about the mechanism of reference if the singular term in question is a proper name such as 'Aristotle' or 'the Big Bang'. We will consider two theories in what follows; the first is an internalist theory, *the description theory*, and the second an externalist theory, *the causal theory*.

According to Russell (1905, 1919), proper names in natural language are abbreviations of definite descriptions. So, for instance, the proper name 'Aristotle' may be an abbreviation of the definite description 'the teacher of Alexander the Great'. Then what I said above about the connection between the definite description 'the teacher of Alexander the Great' and Aristotle carries over to the relation between the proper name 'Aristotle' and Aristotle; namely, what bridges the gap between the proper name 'Aristotle' and the man Aristotle is our understanding of 'the teacher of the Alexander the Great' and our ability to pick out the appropriate object.<sup>2</sup> Kripke (though not Russell himself) called this theory *the description theory (of the reference of proper names)*, or *descriptivism* for short. Obviously, the description theory is an internalist theory of reference.

Frege is considered another description theorist. As we saw in Chapter 2, Frege thought that proper names had senses. Frege says that the sense of 'Aristotle' might be *the disciple of Plato and teacher of Alexander the Great* (even though he concludes that the sense of a proper name in natural language



is often ambiguous). Since this sense can also be expressed by the definite description 'the disciple of Plato and teacher of Alexander the Great', we may as well say that 'Aristotle' is an abbreviation of 'the disciple of Plato and teacher of Alexander the Great'. Generally, proper name *N*'s having sense *S* is no different from *N*'s being an abbreviation of the description '*S*', where '*S*' is the description having sense *S*.

What seems to lend much initial support to the description theory is the fact that when we don't know what a speaker is referring to by some proper name, we just ask him what he is referring to, and he will often give us a definite description of the denotation. For instance, suppose my colleague says, 'John is a very good student'. Then I ask, 'Which John are you talking about?', and he may say, 'The philosophy major who took your philosophy of language course last semester'. Ah, I got it. Like in this case, the speaker can often tell you what description he associates with the proper name he is using.

One problem with the original Frege–Russell description theory, however, is that it is unclear what description ought to be associated with each proper name. Two people may use the same proper name but associate different definite descriptions; for instance, with 'Aristotle' Adam may associate the description 'the teacher of Alexander the Great' and Betty may associate the description 'the most famous disciple of Plato'. Then who is correct in his/her understanding of 'Aristotle'?

In order to solve this problem, John Searle (1958) offered a revision to the original Frege–Russell description theory. His version of the description theory is called *the cluster concept theory*. According to Searle, the proper name 'Aristotle' is an abbreviation of not a single definite description but a cluster – a disjunction – of a sufficient but vague and unspecified number of descriptions. So 'Aristotle' is an abbreviation of, for instance, 'either the teacher of Alexander the Great, or the most famous disciple of Plato, or the author of *Metaphysics*, or the most influential philosopher in antiquity, or ...'. So Adam and Betty are both correct in their understanding of the name 'Aristotle'. Searle argues that this vagueness involved in proper names is what gives proper names usefulness distinct from that of definite descriptions. We can publicly refer to and discuss one and the same object without being bogged down about its identity conditions; for instance, we can refer to and discuss Aristotle without determining whether he is conceived as the teacher of Alexander the Great or the most famous disciple of Plato.

### 5.3 Kripke's Criticism of the Description Theory

Kripke (1980) argued against the description theory. He had at least three lines of argument: the modal, epistemological, and semantic arguments. His arguments are applicable to Searle's cluster concept theory, too, but for the

sake of simplicity, we will consider them as applied to the Frege-Russell original description theory.

In his modal argument, Kripke contends that a proper name cannot be an abbreviation of a definite description because a proper name is a rigid designator, denoting one and the same object in all possible worlds in which the object exists; in contrast, definite descriptions are non-rigid, denoting different objects in different possible worlds. For instance, 'Aristotle' denotes Aristotle in all possible worlds in which Aristotle exists, but 'the teacher of Alexander the Great' denotes whoever is the teacher of Alexander the Great in each possible world, e.g., Socrates or Plato (Sections 2.7 and 2.8). So 'Aristotle' cannot be an abbreviation of 'the teacher of Alexander the Great'.

The epistemological argument is somewhat analogous to the modal argument. Using the same example about Aristotle, the argument goes like this: 'Aristotle' cannot mean the same thing as 'the teacher of Alexander the Great', for if it did, then the sentence 'Aristotle is the teacher of Alexander the Great' must be an analytic sentence, thus expressing an *a priori* proposition. But it does not; we learned that Aristotle is the teacher of Alexander the Great *a posteriori*. To generalize, no proper name '*N*' can mean the same thing as a definite description 'the *D*' because '*N* is (identical with) the *D*' always expresses an *a posteriori* truth.

Kripke's semantic argument simply points out the fact that, especially in the case of the names of historical or famous figures, speakers often do not have any definite description in mind that picks out the proper denotations. Sometimes they have correct but incomplete indefinite descriptions that fail to single out the proper denotations; other times they have incorrect definite descriptions; yet other times, they have definite descriptions that are ultimately circular and fail to fix proper denotations. Still, they are taken as referring to proper denotations.

For instance, a lay person may be taken as referring to the famous physicist Richard Feynman by his use of the proper name 'Richard Feynman', but if we ask him who Richard Feynman is, his answer may be just 'a physicist or something', which does not single out Richard Feynman from other physicists. Lots of people can refer to Albert Einstein by their use of the proper name 'Einstein', but if we ask them who Albert Einstein is, they often say 'the inventor of the atomic bomb', even though Einstein did not invent the atomic bomb. (The atomic bomb was created by a group of scientists led by Robert Oppenheimer in the Manhattan project.)

In a famous fictional example Kripke offers, a man named 'Schmidt' in fact proves the incompleteness of arithmetic but dies, and Gödel<sup>3</sup> gets hold of the proof and publishes it in his name and becomes famous. I associate with the name 'Gödel' the description 'the man who discovered the incompleteness theorem'; still, our intuition tells us that when I say 'Gödel', I am referring to Gödel and not Schmidt even though that description applies to Schmidt and not Gödel.<sup>4</sup>

Some more knowledgeable people may say Albert Einstein is ‘the man who discovered the Theory of Relativity’. Unlike ‘the inventor of the atomic bomb’, this description is a correct description of Einstein, so it should ground the denotation of ‘Albert Einstein’ to the physicist Albert Einstein, right? Well, the problem is that this description contains another proper name, ‘the Theory of Relativity’. According to the descriptivist picture, you can pick out the correct physicist Albert Einstein by that description only if you can pick out the correct theory by ‘the Theory of Relativity’. However, many people, if asked what the Theory of Relativity is, can only say, ‘Einstein’s theory’. So, to ground the denotation of ‘Einstein’, we need to ground the denotation of ‘the Theory of Relativity’, but to ground the denotation of the latter, we need to ground the denotation of ‘Einstein’ – obviously this is circular, and fails to ground the denotation of ‘Einstein’ according to the descriptivist picture. Yet, it seems that lay people are successfully referring to the famous physicist Albert Einstein when they say ‘Einstein’.

Earlier, as support for the description theory, I cited the fact that the users of a proper name can often give you the appropriate definite description of the name’s denotation they have in mind. The semantic argument given above shows, however, that that is not always the case. It is not generally true that people refer to the correct denotations by proper names because they have appropriate definite descriptions in mind.

Kripke concludes that the above three lines of argument, the modal, epistemological, and semantic arguments, show that, contrary to the description theory, proper names cannot be abbreviations of definite descriptions.

## 5.4 The Causal Theory

As an alternative to the description theory, Kripke (along with Donnellan 1970) set forth an externalist theory, *the causal theory (of the reference of proper names)*. Kripke argues that what bridges the relevant proper name and its denotation is the external causal chain from person to person, which starts with the naming of the object and ends with the present uses of the name. In the case of ‘Aristotle’, when Aristotle was born, he was named ‘Aristotle’ (or Ἀριστοτέλης) at the (formal or informal) naming ceremony. Afterwards his parents, relatives, friends, disciples, etc., called him ‘Aristotle’, including Aristotle himself. Aristotle put his name ‘Aristotle’ in his writings. Then the people in the next generation hear or otherwise receive the name from their predecessors (including books) and use the name to refer to Aristotle. Then the next generation, and then the next ..., down to the present. Now you hear or otherwise receive the name from your predecessor, either your philosophy teacher or an author of a philosophy book you read, and, subsequently, you call the Greek philosopher Aristotle ‘Aristotle’. According to Kripke, the proper name ‘Aristotle’ you use denotes Aristotle because

Aristotle is the original source of the causal chain from people to people, generation to generation, that resulted in your use of the name; it is not because you associate an appropriate description or concept of Aristotle with the name when you use it. You may not have such a description in mind.

Kripke's causal theory of reference is an instance of semantic externalism, which claims, generally, that the meaning or denotation of an expression is determined not by what's in the speaker's mind, but by the causal/social/historical environment of the speaker. Toward the end of the twentieth century and in the early twenty-first century, semantic externalism was all the rage in the philosophical scene. It still has many, many advocates today.

## 5.5 Searle's Defense of the Description Theory

However, there have also been philosophers unconvinced by the externalist theory of meaning. In fact, the number of those people seems to be increasing in recent years. For instance, Searle (1983, Chapter 9), whose cluster concept theory Kripke had criticized earlier, was not at all persuaded by the causal theory. He gave a brilliant defense of the description theory to show that neither Kripke's modal nor semantic argument works. In what follows, we will change the order and examine Searle's response to Kripke's semantic argument first, and his modal argument second. Searle himself did not give an explicit response to the epistemological argument; however, his response to the modal argument carries over to the epistemological argument in a straightforward fashion, as you will see at the end of the section.

In his defense, Searle first points out that Kripke's causal theory is not a *pure* (or total) externalist theory. A pure externalist theory would deny the relevance of what's in the speaker's mind entirely, but Kripke admits that what the speaker has in mind still partially matters. For instance, you may hear the name 'Napoleon' from your history teacher for the first time, and decide to name your pet aardvark 'Napoleon'. Afterwards, your use of the name 'Napoleon' in many contexts will denote the pet, not the famous French general, even though the beginning of the causal chain is still the French general Napoleon Bonaparte. Kripke's explanation of this is that in the usual cases, the new speaker has the intention of denoting by the relevant name the same thing as his predecessor did, whereas in the 'Napoleon' case such an intention is absent. So, Kripke concludes, in the standard cases the causal chain must always be accompanied with the new speaker's intention to denote the same thing as his predecessor. And this intention is in the speaker's mind.

This intention, however, makes the external causal chain redundant and unnecessary, Searle argues; what's in the speaker's mind alone connects the speaker to the right predecessors and, ultimately, the right denotation. Suppose that by 'Aristotle', you mean 'the man (or thing) my predecessor

(e.g., my philosophy teacher) refers to by “Aristotle”. This is a definite description. And if your philosophy teacher successfully refers to the correct philosopher Aristotle by ‘Aristotle’, you will also succeed in referring to Aristotle by ‘Aristotle’, thanks to this description. You piggyback on your predecessor to make a successful reference. This mechanism of reference is sometimes called ‘reference-borrowing’ or ‘parasitic reference’. Obviously, in order for me to succeed in parasitically referring to Aristotle, the chain of parasitic reference must end somewhere, where the speaker succeeds in referring to Aristotle thanks to the non-parasitic description he has in his mind; but that’s not an unlikely scenario. Similarly, one can refer to Richard Feynman by having the definite description ‘the physicist others refer to by “Richard Feynman”’ so long as many others successfully refer to Richard Feynman in one way or another (parasitically or non-parasitically).

As we saw, Kripke says that lay speakers often do not have an appropriate definite description of the denotation when they succeed in referring to it. But when he says this, he seems to be overlooking the existence of parasitic descriptions. If we include parasitic descriptions, appropriate definite descriptions are regularly available when we use proper names. Furthermore, if the chain of parasitic relations does not coincide with the Kripkean causal chain, as it does not in the ‘Napoleon’ case mentioned above, it is the parasitic chain, not the causal chain, that determines the reference relation. In that example, the causal chain originates in the French general Napoleon, but the absence of parasitic description enables you to refer to your pet aardvaak. In sum, Searle concludes, Kripke’s semantic argument did not defeat the description theory.<sup>5</sup>

In this context, I should mention another line of defense for descriptivism, called *causal descriptivism* (Kroon 1987). It is a different argument, but with a similar spirit. Causal descriptivism holds that a proper name is the abbreviation of the description of the kind of causal chain Kripke finds necessary. So, to simplify, ‘Aristotle’ is the abbreviation of the description ‘the ultimate source of the causal chain that has led to the current speaker’s use of the word “Aristotle”’. As a general strategy, whenever the externalist finds something, say X, as a necessary external element for the correct reference, the descriptivist can include the description of X as a part of the description necessary for the correct reference. Isn’t it a clever response? Furthermore, it is by no means an ad hoc response: Don’t we always think that the reference of a name we use ought to be its causal source (except for cases like that involving ‘Napoleon’)?

Now, moving on, how about Kripke’s modal argument? It does not fare any better. Recall, Kripke’s point of the argument was that proper names cannot be abbreviations of definite descriptions because proper names are rigid designators whereas definite descriptions are non-rigid. However, first, it’s not as if definite descriptions were all non-rigid by nature; there are

definite descriptions which are rigid, as we saw in Section 3.7; specifically, definite description describing the haecceities (or identity conditions) of individuals are rigid. Also, mathematical truths are true in all possible worlds and the nature of mathematical objects does not change from world to world; so, for instance, definite descriptions such as ‘the smallest prime number’ and ‘the sum of 1 and 1’ are all rigid designators that rigidly designate number 2.

But second and most important, there is a simple and uniform way to *rigidify* any non-rigid definite description (i.e., to turn it into a rigid designator): just add the phrase ‘in the actual world’ at the end of the original description.<sup>6</sup> So, ‘the teacher of Alexander the Great’ will become ‘the teacher of Alexander the Great *in the actual world*’. Think about it: ‘the teacher of Alexander the Great in the actual world’ is a rigid designator and denotes Aristotle in any possible world in which Aristotle exists. And those who might have said that ‘Aristotle’ was the abbreviation of ‘the teacher of Alexander the Great’ probably would have meant ‘the teacher of Alexander the Great *in the actual world*’. So Kripke’s modal argument does not hold up under scrutiny, either. Searle proposes to include rigidified and parasitic descriptions in the relevant cluster to sustain his cluster concept theory.

Even though Searle did not explicitly examine Kripke’s epistemological argument, his conclusion here can easily carry over to defuse the epistemological argument, too. Recall, the epistemological argument argues that no proper name ‘*N*’ can mean the same thing as a definite description ‘the *D*’ because the sentence ‘*N* is the *D*’ always expresses an a posteriori proposition. However, consider the sentence:

- Aristotle is the teacher of Alexander the Great in the actual world.

This is a true identity statement which has rigid designators on both sides (just like ‘Hesperus is Phosphorus’). So it expresses the necessary truth, which can be known a priori (Section 3.12; again, a different proposition, the proposition that the sentence in question indeed expresses the necessary truth, can be known only a posteriori). Therefore, Kripke’s contention that no proper name ‘*N*’ can mean the same thing as a definite description ‘the *D*’ because the sentence ‘*N* is the *D*’ always expresses an a posteriori proposition, does not hold up.

As a whole this is a very clever line of defense for the description theory. Because of this defense and the more sophisticated version of the description theory that has emerged from it, the internalism–externalism debate about the reference of proper names is now in limbo. While there is no doubt that externalism is still the more popular view, the jury is still out as to whether internalism is correct or externalism is correct about the reference of proper names.

## 5.6 The Meaning of Natural Kind Terms; 'Twin Earth'

*Natural kinds* are kinds of things grouped together by nature, as opposed to human convention. *Natural kind terms* are names of natural kinds. For instance, 'Homo sapiens', 'dog', 'cat', 'tiger', 'water', 'gold', 'silver', 'aluminum', 'photon', and 'electron' are natural kind terms; 'table', 'chair', 'house', and 'automobile' are not. Each natural kind has a common underlying physical-chemical-biological structure as an essential property. For instance, water is  $H_2O$  and gold is chemical element Au with atomic number 79. In contrast, a non-natural kind like table has no such structure or essential property; we can modify or expand the concept of *table* in any way suitable. What counts as a table is largely a conventional matter.

Internalism about the meaning of natural kind terms is the view that the meaning of natural kind terms is determined solely by what's in the speaker's mind, whether it is descriptions or something else, such as the perceptual experience. Kripke and Hilary Putnam argued that internalism about the meaning of natural kind terms is incorrect. In particular, Putnam set forth the famous 'Twin Earth' thought experiment to show that two speakers who have identical mental states may mean different things by the same natural kind term.

In that thought experiment, Putnam (1975) asks us to imagine a planet in another galaxy, called 'Twin Earth'. As the name may suggest, Twin Earth is nearly an exact duplicate of Earth, having exactly the same mountains, oceans, cities, buildings, people, etc., in times and places corresponding to those on Earth. Aristotle was born in 384 B.C. in Stagira, Greece, but moved to and lived in Athens until 347 B.C., when he left Athens to become the teacher of Alexander the Great, and died in 322 B.C. Likewise, Twin Aristotle was born in 384 B.C., Earth time, in Twin Stagira, Twin Greece, and moved to and lived in Twin Athens until 347 B.C., Earth time, when he left Twin Athens to become the teacher of Twin Alexander, and died in 322 B.C., Earth time. You get the idea.

I said, 'Twin Earth is *nearly* an exact duplicate of Earth'. There is one crucial difference: where there is water on Earth, there is another substance, Twin water, or *t'*water for short, on Twin Earth. While the chemical composition of water is  $H_2O$ , the chemical composition of *t'*water is completely different, say XYZ. Oceans, lakes, and rivers on Twin Earth are filled with XYZ instead of  $H_2O$ , XYZ falls from the sky when the barometer falls, Twin Earthlings drink XYZ instead of  $H_2O$ , etc. Water does not exist on Twin Earth, and *t'*water does not exist on Earth.

Now, suppose the year is 1750, when the nature of water (and thus, *t'*water as well) has not yet been discovered and there is no interaction between Earth and Twin Earth. Suppose the Earthling Oscar says something about water, say 'Water is wet'. Of course, in exactly the same time Twin Oscar, or

T'Oscar for short, says the same thing on Twin Earth, 'Water is wet'. The question Putnam raises is: what does Oscar mean by 'water', and what does T'Oscar mean by 'water'?

Putnam argues that in this situation, Oscar means water, or  $H_2O$ , by 'water' while T'Oscar means t'water, or XYZ, by 'water'. This is because Oscar is surrounded by water and has causal contact only with water and not with t'water, whereas T'Oscar is surrounded by t'water and has causal contact only with t'water. But note that Oscar's and T'Oscar's physical and mental states are exactly the same.<sup>7</sup> The difference in the meaning of 'water' as Oscar and T'Oscar use the word is entirely due to the difference in their physical environment and not in their experiences. Therefore, as Putnam quips, "the meanings just ain't in the head," i.e., semantic internalism is incorrect.

*Narrow content* is the kind of mental content that does not depend on the subject's environment. *Wide content* is mental content that is not narrow, i.e., the kind of mental content that is affected by the subject's environment. In these terms, Putnam is saying that our mental content at least about natural kinds is wide; Oscar and T'Oscar have different wide mental contents even though their internal experiences are identical.

This is the famous 'Twin Earth' argument against internalism about the meaning of natural kind terms. Many philosophers, who were convinced by Kripke's causal theory of reference, were also convinced by this argument (and its variations like Burge's (1979)) that externalism is correct also for the meaning of natural kind terms.

## 5.7 Two Internalist Responses

However, there are also many philosophers who are not convinced by this argument. There have been various objections, but here I will focus only on two responses: the first response insists that Oscar and T'Oscar have the same mental content; the second grants that Oscar and T'Oscar have different mental contents, but contends that this can be accounted for even from an internalist point of view.

The first response insists that, contrary to Putnam's contention, Oscar and T'Oscar have the same mental content. Nobody on Earth or Twin Earth can tell the difference between water and t'water in 1750. So nobody can be thinking of just water or t'water. Accordingly, 'water', either used on Earth or Twin Earth, cannot be referring to just water or just t'water. In particular, Oscar and T'Oscar are not referring to different substances, water and t'water respectively, by the same word 'water'; they are both referring to *water or t'water*. Some, though not many, of my students confess that they have this intuition. Since Putnam's thought experiment relies only on our raw intuition, this response should be as good as the one Putnam counts on.



In fact, Putnam, to his credit, mentions a real-life example in which two different chemical compounds are given a single name: jadeite is a silicate of sodium and aluminum, and nephrite is a silicate of calcium and magnesium, but they are grouped together and given the common name 'jade' in English because to lay people they are indistinguishable. (Their chemical difference was discovered only in the second half of the nineteenth century.) We can even construct a (non-actual) scenario in which nephrite exists only in the Americas, and that people in the Old World encountered only jadeite before the discovery of the New World; still it is easily imaginable that they have come to call both jadeite and nephrite 'jade'. If this can happen for jadeite and nephrite,<sup>8</sup> how can the same not happen for water and t'water?

Furthermore, as I said, Putnam's Twin Earth argument relies on our intuition, but our intuition seems to waver depending on how his story unfolds. Suppose, for instance, that Earth and Twin Earth discover each other and Earthlings and Twin Earthlings begin traveling to the other planet before they discover the nature of, and the difference between, water and t'water. In that case, they will of course call both water and t'water 'water' initially, and it is likely that this usage will stick even after the discovery of the difference, just like in the jade case. On the other hand, if the discovery of the nature of water and t'water precedes the interaction between Earth and Twin Earth, it seems less likely that both water and t'water will end up being called 'water'. This may make one wonder if the word 'water' already had determinate meaning(s) when Oscar and T'Oscar uttered the word.

Of course, if our use of 'water' denotes both water and t'water, 'water' will not be a natural kind term. But there is no problem with that. When we initially assumed that 'water' was a natural kind term, we were assuming that nothing like t'water existed. If we are asked to assume that t'water exists, we must suspend our assumption, at least momentarily, that 'water' is a natural kind term.

So that's the first internalist response to Putnam. I will be brief about the second response I have in mind. Searle (1983, Chapter 8) was sympathetic to the first response I have just given, but also gave this response. He grants, for the sake of argument, that Oscar and T'Oscar have different mental contents when they see water/t'water and say 'Water is wet'; but he contends that this can be explained easily from an internalist viewpoint. By 'water', both Oscar and T'Oscar mean something like: 'the mereological sum of things which have the same chemical composition as the stuff which is causing the present visual experience in the current speaker's mind'. Oscar and T'Oscar have two tokens of the same description in mind, but the tokens denote numerically different visual experiences in numerically different minds, and they are caused by different substances, water and t'water; as a result, the tokens obtain different contents. Searle maintains that this is perfectly consistent with the internalist picture.

This response has an undeniable similarity to the causal descriptivism I mentioned in Section 5.5. Both try to defend descriptivism by including the relevant causal relation and causal source in the description needed to fix the reference. In the present case, the different causal sources result in Oscar's and T'Oscar's different mental contents through the relevant description. If Searle is correct, then the simple equations often assumed, internalism = (the advocacy of) narrow content and externalism = wide content, need to be reconsidered.

Again, these are just two of many responses philosophers have given to Putnam's argument. Just as in the case of proper names, there is no easy way to determine whether internalism is correct or externalism is correct with regards to the meaning of natural kind terms. Generally, thought experiments are fun, but they at best reveal our raw, uneducated intuitions. They rarely tell us if those intuitions are based on the correct theories of the matter or what those correct theories are.

To summarize Sections 5.2–5.7, despite the enormous popularity of semantic externalism at the end of the twentieth century to the early twenty-first century, it is fair to say that internalism is far from being defeated; it is still an open question which of the two positions is the correct position about meaning and content. Or, perhaps, a third position that synthesizes the two may emerge in the future.

## 5.8 Narrow vs Wide Content; Intrinsic vs Extrinsic Properties

At the end of Section 5.6 I introduced the distinction between *narrow* and *wide content*. In this section, I would like to introduce another distinction related to it: *intrinsic* vs *extrinsic properties*. To conclude Part A of this chapter, in the next section I will introduce the concept of *supervenience*, which is also related to these distinctions.

How are these concepts related to one another? The following account of narrow and wide content should give you an idea:

- Narrow content and wide content
  - *Narrow content* is the kind of mental content that does not depend on the subject's environment. Having narrow content (that *P*) is an *intrinsic property* of the subject. Narrow content *supervenes on* the subject's mental states.
  - *Wide content* is mental content that is not narrow, i.e., the kind of content that depends on the subject's environment. Having wide content (that *P*) is an *extrinsic property* of the subject. Wide content *does not* supervene on the subject's mental states.

What is an intrinsic property? What is an extrinsic property? They are defined as follows:

- Intrinsic property and extrinsic property
  - An *intrinsic property* of object  $x$  is a property  $x$  has in itself, regardless of  $x$ 's environment. E.g., mass, the property *having blue (brown, green, etc.) eyes*, the property *being human*, etc.
  - An *extrinsic property* of object  $x$  is a property  $x$  has at least partly because of  $x$ 's environment. E.g., weight, the property *being a sibling (brother, sister, son, daughter, father, mother, etc.)*, the property *being an American citizen*, etc.

For instance, compare the mass and weight of a physical object like us. Our weight changes when we are on the moon; we are six times as weighty on Earth as we are on the moon. That's because our weight is not our intrinsic property but is dependent partly on the gravitational field around us. In contrast, our mass is constant whether we are on Earth or the moon. So mass is our intrinsic property. Of course, this does not mean that we don't change our mass; we do, by eating more and gaining more muscles or fat, for instance. But our mass changes only when something inside us changes.

Similarly, our eye colors are our intrinsic properties, whereas our properties *being a sibling (brother, sister, etc.)* are extrinsic. I can lose the property *being a sibling* if my only sibling (brother or sister) dies; then I no longer will be a sibling even though nothing changes in my body.

Earlier in Section 1.4, we discussed  $n$ -place relations and treated properties as 1-place relations. By filling in  $n-1$  slots of an  $n$ -place relation, we can turn the  $n$ -place relation into a property. The resulting property may be called a *relational property*. So, for instance, if we have a 2-place relation  $x$ 's *being a sibling of*  $y$  or  $x$ 's *being taller than*  $y$  and fill in Adam for  $y$ , the resulting property  $x$ 's *being a sibling of Adam* or  $x$ 's *being taller than Adam* will be a relational property.

As you can easily imagine, relational properties are often extrinsic properties, for something's having a certain relational property or not depends not only on that thing itself but also the other things involved in the property. I may lose the property *being taller than Adam* just because Adam grows taller. There are, however, relational properties that are intrinsic; for instance, the relational property *thinking about the Fountain of Youth* seems to be an intrinsic property determined by our mental states alone.

If narrow content exists, then *having narrow content that*  $P$  (for some specific proposition  $P$ ) will be an intrinsic property, for whether we have this content or not is determined solely by our mental/bodily states and not by our environment. If wide content exists, then *having wide content that*  $P$  will be our extrinsic property, for whether we have that content or not is determined

partly by our environment. For instance, Oscar's having wide belief content that water is wet is determined partly by his environment, where there is much water and no t'water. T'Oscar may be in the same mental/bodily state but instead believes that t'water is wet.

One of the most fundamental concepts in ethics and value theory in general is that of *intrinsic value*. Intrinsic value is contrasted with *extrinsic value*. Since value of an object is a property of the object in a broad sense, intrinsic/extrinsic value of an object is a property of the object, too. So:

- Intrinsic value and extrinsic value
  - *Intrinsic value* of  $x$  is value  $x$  has in itself, regardless of  $x$ 's environment.
  - *Extrinsic value* of  $x$  is value  $x$  has partly because of  $x$ 's environment (often connected with intrinsic value of itself or other things).

The things that are most often said to have intrinsic value, mentioned ever since the time of Plato, are pleasure and pain. Plato maintained that pleasure is good in itself and that pain is bad in itself; the only reason some 'excessive' pleasure seems bad is because it tends to bring less pleasure and more pain in the long run. In contrast, diamonds have only extrinsic value. They have value only because they bring us much pleasure. They would lose most of their value if the social, economical, or geological environment were different (e.g., if diamonds became readily available and/or we stopped valuing them) and they stopped bringing us pleasure. So-called 'instrumental' value is a typical instance of extrinsic value, value things have because they are instrumental to the production of intrinsic value. For example, money has instrumental value to often bring intrinsic value through pleasure or happiness.

Again, intrinsic and extrinsic value are instances of intrinsic and extrinsic properties. This does not mean that intrinsic and extrinsic value do indeed exist, for, among other things, value in general may not exist; moral anti-realism (Sections 4.7 and 4.9) and, more generally, value anti-realism, may turn out to be correct. However, we will not pursue that issue any further.

## 5.9 Supervenience

Another way to describe the difference between narrow content and wide content is that narrow content *supervenies (or is supervenient) on* the subject's mental states while wide content does not.

Supervenience, the concept made popular by the American philosopher Donald Davidson in the 1970s, is a curious concept. It is widely used in many areas of philosophy, but not used at all in any discipline outside of philosophy. Broadly speaking, supervenience is a dependence relation between two kinds of properties. More precisely, it can be defined as follows:

- Supervenience  
Set *A* of properties *supervenes* (or *is supervenient*) on set *B* of properties iff any objects that are indistinguishable with respect to the *B*-properties must be indistinguishable with respect to the *A*-properties (or, *B*-twins must be *A*-twins).

Sets *A* and *B* here are called the *supervenient set* and the *base* (or *subvenient*) set (or *supervenient base*), respectively. This definition is summarily expressed by the slogan, 'There cannot be an *A*-difference without a *B*-difference'.

To get a better idea about supervenience, let's consider the following five claims:

- (a) Moral properties supervene on natural (or non-moral) properties.
- (b) The macro-level properties in the physical world supervene on its microphysical properties.
- (c) Mental properties supervene on physical properties.
- (d) The mental properties of a subject supervene on his/her physical properties.
- (e) Modal properties of an object supervene on its non-modal (or categorical) properties.

(a), the supervenience of moral properties on non-moral properties, seems like a plausible claim in ethics. Even if moral realism is correct (Section 4.7) and there are moral properties existing independently of non-moral properties such as physical properties, it seems impossible for the moral properties to be distributed in different ways in two physically identical worlds. For instance, it seems impossible for abortion to be morally wrong in one of those worlds and morally permissible in the other world. Similarly, (b), the supervenience of the macrophysical properties on the microphysical properties, seems like a plausible thesis. For instance, the shapes of medium-sized physical objects around us, like chairs and tables, cannot be different unless the locations of some of the particles involved are different.

Davidson (1970) originally introduced the concept of supervenience in his discussion on the mind-body problem, and made claim (c), the supervenience of the mental on the physical. Davidson embraced property dualism (Section 4.5) and maintained that even though every object in the world is material, there are two kinds of properties, mental properties and physical properties. For example, no single thing but many different kinds of things are going on in my brain when I think about the Fountain of Youth at different times. Different kinds of things are going on in my brain and in your brain when I think about the Fountain of Youth and when you think about the Fountain of Youth. So the mental property *thinking about the Fountain of Youth* cannot be identified with any single physical property. Still, mental properties are not independent of physical properties since they are

supervenient on physical properties: two physically indistinguishable worlds cannot have different distributions of mental properties.

From this consideration Davidson came to uphold the view that combines substance physicalism with property dualism. Later critics pointed out, however, that the supervenience of the mental on the physical is also consistent with much stronger forms of dualism such as substance dualism. Going back to claim (a) above, the supervenience of moral properties on natural properties does not imply that moral properties are nothing but natural properties. For instance, the (alleged) wrongness of abortion may not be identifiable with any physical feature of abortion. Another, perhaps more reasonable, explanation of the supervenience may be that the wrongness of abortion is distinct from, but *correlated* with, physical features of abortion; so two physically identical situations cannot be different with respect to the wrongness of abortion. Analogously, the (alleged) supervenience of the mental on the physical does not imply that the mental is nothing but the physical. It may just be that the mental is correlated with the physical in such a way that two physically identical situations cannot be different with respect to the mental. Davidson and his followers took ‘*A* is supervenient on *B*’ to mean ‘*A* is nothing but *B*’, but this identification is untenable.

(d), the claim that our mental properties supervene on our physical properties, can be true if our mental content is narrow but is false if our mental content is wide. The physical twins Oscar and T’Oscar can have different mental contents if mental content is wide. Oscar can be thinking that water is wet, while T’Oscar can be thinking that t’water is wet. Note the difference between (c) and (d): (c) is about the supervenience of the mental world *as a whole* on the physical world *as a whole*, whereas (d) is about the supervenience of the mental properties *of an individual* on the physical properties of the individual. Supervenience of the former kind is called *global supervenience*, while supervenience of the latter kind is called *local (or individualistic) supervenience*. The existence of wide mental content would falsify (d) but not (c) because if mental content is wide, Oscar and T’Oscar are two physical twins that have different mental contents, but Oscar’s overall environment and T’Oscar’s overall environment are not two physical twins.

Finally, some philosophers argue for (e), the supervenience of modal properties on non-modal (or what some call ‘categorical’) properties,<sup>9</sup> and use it as the basis for the problem of grounding raised against the distinctness between Goliath and Lump1 (Section 3.8). They maintain that two things identical with respect to non-modal properties cannot have different modal properties. Since Goliath and Lump1 do have the same non-modal properties and are indistinguishable in the actual world, they cannot have different modal properties such as *x can survive the loss of one hand*. Therefore, Goliath and Lump1 must be one and the same thing. In response, some advocates of the constitution view contend that while Goliath and Lump1 may have the same *intrinsic* non-modal properties, they can and indeed do have different

*extrinsic* non-modal properties. For instance, Goliath does but Lump1 does not have such extrinsic non-modal properties as *x is an artifact*, *x is a statue*, and *x is a piece of art*, which arise from the way people treat the statue. Therefore, they argue, (e) may be true but does not lead to the problem of grounding.

## PART B: INTERNALISM AND EXTERNALISM IN EPISTEMOLOGY

### 5.10 The JTB Theory of Knowledge

In the first three sections of Part B, we will consider an internalism–externalism debate in epistemology about epistemic justification, i.e., whether a subject’s being justified in having a certain belief is determined solely by what’s in the subject’s mind or not. After laying a sufficient background in this section, I will present two internalist theories of epistemic justification, i.e., foundationalism and coherentism, and one externalist theory, reliabilism, in the next two sections. The fourth section of Part B, Section 5.13, examines Putnam’s externalist argument against skepticism, in which Putnam uses his semantic externalism in an attempt at refuting the ‘brain in a vat’ skepticism.

Epistemology is also called *theory of knowledge*. What is knowledge? It is often said that there are at least two kinds of knowledge: *know-how* and *know-that*.<sup>10</sup> Know-that, also called *propositional knowledge*, has the form ‘Subject *S* knows (or has knowledge) that *P*’, where *P* is a proposition. I know that Aristotle is the teacher of Alexander the Great. You know that Mt Everest is the highest mountain in the world. These are examples of know-that. At the same time, I know how to swim, and how to ride a bike; I don’t know how to play guitar, or how to drive a stick shift. These are examples of know-how. Know-how is a skill. It is often said that know-how may involve some propositional knowledge, but is not reducible to it. I may know how to ride a bike, but may not have propositional knowledge about how to keep a bike standing at a slow speed. A mechanical engineer may have that propositional knowledge but still may not be able to bike. Most epistemologists are interested only in propositional knowledge, so that will be the subject of our subsequent discussion.

In this book we uphold the historically most prominent definition of propositional knowledge, *the JTB theory of knowledge*. According to the JTB theory, knowledge is Justified True Beliefs. The JTB theory is so called because of that. More precisely:

- JTB theory of knowledge  
Subject *S* knows that *P* iff

- (a) *S* believes that *P*;
- (b) *P* is true; and
- (c) *S* is justified in believing that *P*.

This definition of knowledge originates in (who else?) Plato.

Among the three necessary conditions for knowledge, the need for (a) and (b) is universally accepted: in order for *S* to know *P*, *S* must believe *P*, and *P* must be true. Knowledge is a species of beliefs, and a falsehood cannot be the content of knowledge. Historically the problem has always been about (c): for your belief to be knowledge, it must be not only true but justified. Condition (c) is necessary to exclude lucky guesses from knowledge. Suppose I guessed that the horse named 'Justify' would win Kentucky Derby just because it had a great name, and suppose the horse in fact won.<sup>11</sup> So I had a belief which turned out to be true, but nobody says that I *knew* Justify would win. In contrast, if I *saw* Justify actually crossing the finish line first, then my true belief afterwards that Justify won may be considered knowledge. So in order to know *P*, I must have a good reason to believe that *P*, or my belief *P* must be *justified* (or *warranted*).

There has been much debate about whether conditions (a) to (c) are collectively the sufficient condition for knowledge. Many people take the famous *Gettier problem* (after Edmund Gettier 1963) to show they are not. In one of the two counterexamples he gives, Gettier asks us to think of a situation in which Smith and Jones are competing for a job, and Smith comes to believe the following propositions:

- (d) Jones is the man who will get the job, and Jones has ten coins in his pocket;
- (e) The man who will get the job has ten coins in his pocket.

Smith is justified in believing (d) because he heard the company president say that Jones would get the job, and also because he saw ten coins in Jones's pocket. From (d) Smith logically infers (e), so he is justified in believing (e), too. A surprising twist in this story is that it turns out that it is Smith, not Jones, who actually gets the job, and that, unknown to Smith, he also has ten coins in his pocket. So (e) turns out to be true. Again, Smith believed, and was justified in believing, (e). So, if we follow the letters of the JTB theory, Smith must have knowledge that (e) is true. But that's contrary to our view that Smith was just lucky to have a belief that turns out to be true. So conditions (a) to (c) cannot be the sufficient condition for knowledge, Gettier argued. Because of this argument, many philosophers have given up the JTB theory.

Interesting as it may be, we are not going to discuss this argument in this book (sorry!) because it is a little away from the main topics of the book. Instead, I would like us to think more about what 'justified' ought



to mean in the above definition of knowledge. In particular, I am going to introduce two internalist theories of epistemic justification, foundationalism and coherentism, and one externalist theory, reliabilism, which give very different interpretations of ‘justified’.

### 5.1.1 Internalist Theories: Foundationalism and Coherentism

Foundationalism and coherentism are the only well-known internalist theories of epistemic justification. Foundationalism is by far the more popular. The structures of justification they set forth are figuratively called ‘the pyramid’ and ‘the raft’, respectively, even though I think ‘the (spider) web’ is a better name for the latter. We will see why they are so called very shortly.

According to foundationalism, there are two kinds of beliefs: *basic (or foundational) beliefs* and *non-basic beliefs*. The basic beliefs are beliefs that require no inferential justification. The non-basic beliefs are beliefs that ought to be justified inferentially by more basic beliefs, i.e., some other non-basic beliefs or some basic beliefs. Thus, our system of beliefs has a hierarchical, pyramid-like, structure (Figure 5.1).

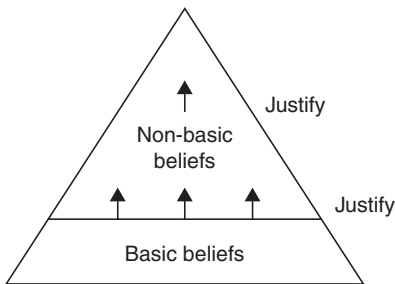


Figure 5.1 The pyramid.

A non-basic belief may be justified by another non-basic belief, which, in turn, may be justified by yet another non-basic belief, and so on; however, this chain cannot go on indefinitely or in circles. Some justified beliefs in the chain must be justified by basic beliefs. So, every justified non-basic belief is justified ultimately by some basic beliefs.

Do the basic beliefs require justification at all? The answer depends on what version of foundationalism one embraces. In one version of foundationalism, the basic beliefs are said to be ‘self-justifying’ or ‘self-evident’. I take ‘self-justifying’ to mean that the basic beliefs require no justification. Descartes seemed to have a similar view, but this version of foundationalism has lost

much popularity. Even if Descartes was right and a few beliefs such as ‘I think’ and ‘I exist’ might be self-evident, there are just not enough beliefs of such status to form a sufficiently broad base for a system of knowledge. (Descartes ‘proved’ the existence and veracity of God and used them to support the extant system of knowledge, but most people find that part of Descartes’s argument quite dubious.)

In another, perhaps more popular, version of foundationalism (Russell 1912; Ayer 1940), the basic beliefs are to be justified non-inferentially by some ‘non-doxastic’ (= non-belief) mental states. (‘Doxa’ means ‘belief’ in Latin.) Let me give you an example to illustrate the point. I am hungry. So I come to believe that I am hungry. Then I come to believe that my stomach (literally) is empty. Here my belief that I am hungry is justified by my mental state of feeling hunger and no beliefs. Recall (Section 4.2) that hunger, arguably, is not a propositional attitude and does not involve a proposition; it is, thus, a non-doxastic state. So the belief that I am hungry is a basic belief justified by the non-doxastic state, the feeling of hunger; and this justification is not by means of an inference from a proposition to another proposition, for hunger does not involve a proposition. The belief that my stomach is empty is justified inferentially by the belief that I am hungry; I can make an inference from the belief that I am hungry to the belief that my stomach is empty as a way of justifying the latter belief. So the belief that my stomach is empty is a non-basic belief justified by the basic belief that I am hungry.

Most philosophers who embrace this version of foundationalism think that even behind perceptual beliefs, there are non-doxastic mental states of sensation. When I see, i.e., obtain a belief by vision, that there is a red book in front of me, there is a non-doxastic sensation of redness that justifies the belief. The raw data we obtain by sensation are called *sense-data*. A perceptual belief is justified if it interprets the relevant sense-data correctly, and it is unjustified if it interprets the sense-data incorrectly. Foundationalists often mention perception, memory, and introspection as instances of mental states involving non-doxastic mental states.

In summary:

- Foundationalism  
S’s belief *B* is justified iff either
  - (a) *B* is a basic belief; or
  - (b) *B* is a non-basic belief ultimately justified inferentially by S’s basic beliefs.

In contrast to foundationalism, coherentism (Lehrer 1974; Bonjour 1985) does not assume any difference in the basicness of beliefs. All beliefs are of equal status. Beliefs are collectively justified so long as they cohere, or fit together. The justification relation among beliefs looks like a raft or a spider web like this (Figure 5.2).

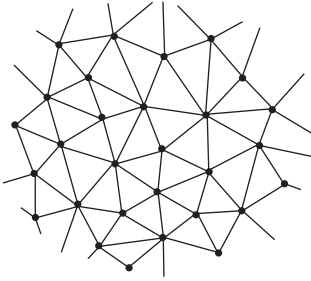


Figure 5.2 The raft (or web).

Generally, beliefs firmly embedded in our belief system are near the center of the web; they are interconnected with many other beliefs and do not change easily. They include beliefs about the universe in general and beliefs about logical and mathematical truths. Beliefs at the periphery of the web are less interconnected with other beliefs and can change more easily. They include beliefs about individual objects. It is just a matter of degree, however, how firmly beliefs are held in the web, and even beliefs near the center may be replaced if necessary.

- Coherentism  
S's belief *B* is justified iff *S* has a coherent set of beliefs which includes *B*.

Coherentism is a holistic theory: beliefs are justified or unjustified as a group, and not individually. Unlike the aforementioned popular version of foundationalism, coherentism puts no particular significance on non-doxastic mental states.

There has been much debate between foundationalists and coherentists about which theory draws a more plausible picture of epistemic justification. Since our goal is not to determine which of the two is better, but to contrast them with an externalist theory of epistemic justification, reliabilism, I will ignore the debate here, except to point out one major problem for foundationalism and two for coherentism.

For foundationalism, the problem is about the non-doxastic mental states. It's not clear whether we have a non-doxastic mental state whenever we obtain a perceptual belief. In fact, that seems unlikely. When we observe our environment, we obtain perceptual beliefs immediately, without any mediation of non-doxastic mental states (Chisholm 1942; Sellars 1956). But then many of our beliefs won't be justified in the foundationalist picture.

Perhaps the most serious problem with coherentism is about the concept of coherence itself: what coherence is has never been made clear. People talk about the coherence of beliefs only in such figurative terms as the beliefs'

‘fitting together’ or ‘hanging together’, and nobody has ever succeeded in giving a precise account of what these mean.

Another major problem for coherentism, even when we operate under the unformalized concept of coherence, is the apparent fact that coherence of beliefs can be achieved too easily. For instance, think of the version of creationism according to which the whole universe was created very recently, perhaps sometime around 4004 B.C. This is a crazy theory, not justifiable in any reasonable sense of justification. However, the defenders of this view make their system of beliefs coherent by adding equally outrageous claims after claims in the system, such as that God buried fossils in the ground when He created the universe, and He did it in order to test our faith in Him, that humans and dinosaurs actually cohabited, that the Bible actually refers to dinosaurs, etc. It is also a well-known fact that some psychiatric patients have outrageous sets of beliefs that are nonetheless internally coherent. So coherence seems to be too weak a condition for epistemic justification.

But, again, it is not our goal here to consider the problems for foundationalism and coherentism or determine which theory is more plausible. Instead, I would like to emphasize the common feature that makes both of them internalist theories. Internalism about epistemic justification holds that what’s in the subject’s mind determines whether a belief is justified or not. Furthermore, if something is in the subject’s mind, it must be accessible to the subject on reflection.<sup>12</sup> On both foundationalism and coherentism, what justifies a belief is other mental states, i.e., other beliefs or non-doxastic mental states. And all those mental states are accessible to the subject on reflection. The subject can simply ask him/herself, ‘Why do I believe this?’, and retrieve other beliefs as its justification. For instance, I can ask myself, ‘Why do I believe that I should eat dinner as soon as possible?’ and retrieve the belief that I am hungry. I can then ask myself again, ‘Why do I believe that I am hungry?’ and retrieve the non-doxastic mental state of hunger. Of course, we can make mistakes. Your psychiatrist may sometimes tell you that you are not retrieving the real justification of your belief. Overall, however, we are better than anybody else at retrieving our own beliefs. This is the feature common to foundationalism and coherentism, and that’s why they are both considered internalist theories.

## 5.12 An Externalist Theory: Reliabilism

Until fairly recently, internalism was the only available theory of epistemic justification; but in the late twentieth century, externalism about epistemic justification emerged as a serious contender. In particular, Alvin Goldman’s (1979) *reliabilism* is now regarded as the most convincing externalist theory that casts doubt on the conventional idea that what justifies one’s beliefs ought to be one’s other mental states.

To motivate reliabilism, let me ask you the following question: What makes my belief that the horse Justify will win unjustified and my belief that Justify won justified? Many people's answer is *not* that the latter belief is supported by other mental states while the former is not. The simple and better answer is that the former belief is produced by pure guessing or wishful thinking while the latter is produced by perception. We all know that perception is generally a reliable belief-forming mechanism while pure guessing and wishful thinking are not. This line of thinking leads us to the following reliabilist definition of epistemic justification:

- Reliabilism  
S's belief *B* is justified iff *B* is produced by a reliable cognitive process.<sup>13</sup>

The following are examples of:

- Reliable cognitive processes  
Perception, memory, deductive reasoning, inductive reasoning from many samples, and introspection.

The following are examples of:

- Unreliable cognitive processes  
Pure guessing (mere hunch), wishful thinking, inductive reasoning from few samples, astrology, palm reading, and the use of the Ouija Board.

A process's being reliable does not mean that the process *always* produces a true belief; it only means that the process *tends* to produce true beliefs. It's a matter of probability. But, surely, if a certain cognitive process tends to produce truth with a high degree, and if your current belief was produced by that process, then there is reason to think that the belief is true; so you are justified in holding that belief. Recall that beliefs' being justified was initially included as a necessary condition for knowledge in the JTB definition in order to eliminate the element of luck. If a true belief was produced by a reliable cognitive process, you were not just lucky to obtain that belief.

Reliabilism is a version of externalism, for subject *S* may not know that the cognitive process that produced belief *B* is in fact reliable; so he may be justified in having belief *B* without knowing that he is justified. In fact, he may never know that he is justified. This probably does not happen in the case of perception, the most typically reliable cognitive process; we all know that perception is reliable. However, think about the following scenario: Suppose that genuine clairvoyance exists as a cognitive process, and that it is in fact a reliable process. Only a few people have the clairvoyance power, so the existence of this power is not yet publicly recognized, and

many of the clairvoyants themselves do not realize that they have that power. (They sometimes obtain some inexplicable visions about certain things, but they do not check their accuracy.) Norman is one of those clairvoyants, who has visions about the US President's whereabouts. Suppose Norman now comes to believe, as a result of such a vision, that the President is currently in the White House. Is he justified in his belief?

The reliabilist says 'yes'.<sup>14</sup> Note, however, that Norman has no way of knowing that his belief is in fact justified, for he has no way of knowing that he has that reliable cognitive mechanism. So what justifies Norman's belief is external to his mind.

Another, more colloquial, expression for 'justification' is 'reason'. But there are two ways in which a belief has a reason: 'Subject *S* has a reason to believe *P*' vs 'there is reason for *S* to believe *P*'. (This is obviously related to the reasons internalism vs externalism that I touched on in Section 5.1.) In these terms, Norman does not have a reason to believe that the President is in the White House, for he does not have an internal justification in his mind; but there is objective reason for Norman to believe that the President is in the White House because there is a reliable cognitive mechanism at work, even though that fact is beyond Norman's reach.

One problem for reliabilism is that the concept of *cognitive process* is unclear. What counts as a single cognitive process? For instance, *vision in broad daylight* is much more reliable than *vision at night*; do they count as different cognitive processes, or is *vision* one cognitive process? Vision is overall still pretty reliable on Earth, but unreliable in some other environments. Then should *vision on Earth* be a different cognitive process than, say, *vision on Saturn*? Similarly, although in the previous lists I listed memory and introspection as reliable processes, they are notoriously unreliable and even deceptive in some circumstances. Shouldn't those cases of memory and introspection count as different cognitive processes? Contrariwise, I listed above inductive reasoning from many samples as a reliable cognitive process and inductive reasoning from few samples as an unreliable cognitive process; but should they really count as different cognitive processes? Shouldn't they instead be combined as one cognitive process, inductive reasoning? This problem is called *the generality problem*. However, just as in the case of foundationalism and coherentism, I will not consider further the problems with reliabilism here.

Instead, I will conclude this section with an examination of the implications of internalism and externalism for the KK Principle, the principle mentioned earlier in Section 3.15. The KK Principle can be expressed in English as follows:

- The KK Principle  
If I know that *P*, then I know that I know that *P*.<sup>15</sup>

Is this principle plausible? It depends. It depends on whether internalism is correct or externalism is correct. Suppose we have the following two principles:

- (P1) If I believe that *P*, then I believe that I believe that *P*.
- (P2) If I am justified in believing that *P*, then I believe that I am justified in believing that *P*.

Then, assuming that the JTB theory of knowledge is correct and that I believe it is, the consequent (the then-part) of the KK Principle will be derivable from its antecedent (the if-part), as follows:

Suppose

- (a) I know that *P*.

Then

- (b) I believe that *P*. (From (a))
- (c) I am justified in believing that *P*. (From (a))
- (d) I believe that I believe that *P*. (From (b) by (P1))
- (e) I believe that I am justified in believing that *P*. (From (c) by (P2))
- (f) I believe that I know that *P*. (From (b), (d), and (e), as I believe in the JTB theory)

But since I obtain this purely by a logical inference,

- (g) I am justified in believing that I know that *P*.

Therefore,

- (h) I know that I know that *P*. (From (a), (f), and (g))

This justifies the KK Principle as a correct principle. So the question is: Are (P1) and (P2) plausible principles?

(P1) seems like a plausible principle. But how about (P2)? This is where whether justification is internal or external will make a difference: if justification is internal and accessible to me, then (P2) is rather plausible,<sup>16</sup> whereas if justification is external to my mind, (P2) has no plausibility at all. Again, according to the externalist conception of justification, I may not know that I am justified. In sum, the KK Principle is a much more plausible principle on the internalist conception of justification than the externalist conception.

### 5.13 Putnam's Semantic Externalist Argument against Skepticism

One of the major topics in epistemology concerns skepticism, especially global skepticism of the kind the 'brain in a vat' hypothesis (Section 4.6) raises. Putnam (1981) tried to refute the 'brain in a vat' skepticism by using his semantic externalism. Since his attempt cuts across the two areas of philosophy we've discussed in this chapter, i.e., semantics and epistemology, I think it is worthwhile to examine it in this section, even though it is still an open question whether semantic externalism is correct.

Assuming something like the causal theory of reference, Putnam argues that I cannot be a brain in a vat. His argument goes like this: Suppose I have been a brain in a vat all my life. And suppose I now entertain the thought that I may be a brain in a vat and even assert 'I am a brain in a vat'. However, the expressions 'I', 'brain', and 'vat' contained in this sentence do not refer to my physical body, physical brain, or physical vats as my utterance of the same expressions would if I were not a brain in a vat and lived in the normal world. For 'I', 'brain', and 'vat', used in the vat-world have never had an appropriate causal contact with my (non-existing) physical body, physical brains, or physical vats. True, in a way they are related to my brain in the vat, but those causal relations don't seem to be the kind of causal relations that make up the reference relations. The references of 'I', 'brain', and 'vat' must, instead, be something like my pictures of myself, brains, and vats in the vat-world (Figure 5.3). Let's call those pictures 'vat-I', 'vat-brains', and 'vat-vats'. However, in the vat-world, vat-I is not a vat-brain in a vat-vat; vat-I is vat-teaching vat-philosophy, having his vat-brain in his vat-body. Thus, the

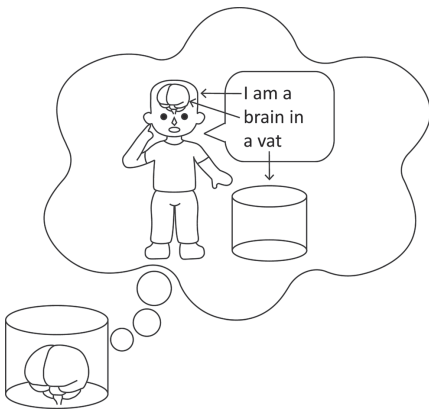


Figure 5.3 'I am a brain in a vat'.



vat-I's assertion of 'I am a brain in a vat' is false. Therefore, I am not a brain in a vat. Since I have now derived from the assumption that I am a brain in a vat the conclusion that I am not a brain in a vat, namely a contradiction, the original assumption that I am a brain in a vat is refuted. QED. (You may know that this form of argument is called *reductio ad absurdum* or *reductio* for short.)

What do you think of this argument? Are you convinced that you are not a brain in a vat? Well, unfortunately, this argument does not work, even if we grant Putnam that semantic externalism is basically correct. There is an error in the argument that has nothing to do with semantic externalism.

In order to examine the argument carefully, we need a little preparation. Let us recall the distinction between the object language and the metalanguage, introduced back in Section 1.7, as we need to distinguish vat-English from normal-English. To repeat, the object language is the language we are talking *about*, and the metalanguage is the language of the narrative, the language we are talking *in*. So if we are talking about vat-English in normal-English, vat-English is the object language and normal-English is the metalanguage.

Oftentimes, the metalanguage and the object language are one and the same language. This happens, for instance, when we are talking about normal-English in normal-English. That's what we have done throughout this book (assuming, of course, that I am not a brain in a vat).

There is an important principle that holds if the object language and the metalanguage are one and the same:

- The Disquotation Principle (for Truth)  
For any sentence  $P$ ,  
' $P$ ' is true iff  $P$ .

Instances of the Disquotation Principle (also called 'the Disquotation Schema' or 'the Equivalence Schema') are: 'Earth is round' is true iff Earth is round, and '2 plus 2 equals 5' is true iff 2 plus 2 equals 5. If the sentence  $P$  contains indexicals, then we need to make a few adjustments; for instance, your assertion yesterday of 'I'm fine today' is true iff you were fine yesterday (not iff I am fine today). However, I won't get into that, as the principle presented above is sufficient for our purposes. 'Disquotation' must be a strange word to you, but this principle is so called because it shows that the predicate ' $x$  is true', the so-called *truth predicate*, in a way works as the inverse operation of quotation: for any sentence  $P$ , if you put it in quotes (i.e., ' $P$ ') and then add the truth predicate at the end (i.e., ' $P$  is true'), it will be equivalent with the original  $P$ .<sup>17</sup> Keep in mind that the Disquotation Principle holds only if the object language is the same as the metalanguage; "'雪は白い' (yuki wa shiroi) is true iff snow is white," though true, is obviously not an instance of the Disquotation Principle.

Furthermore, if we assume, as we do for the sake of simplicity, that all (grammatical) sentences are either true or false, we can have the following falsity version of the Disquotation Principle:

- The Disquotation Principle for Falsity  
For any sentence *P*,  
'*P*' is false iff it is not the case that *P*.

For instance, '2 plus 2 equals 5' is false iff it is not the case that 2 plus 2 equals 5.

Now we are ready to analyze Putnam's argument against the 'brain in a vat' hypothesis. His argument basically goes as follows:

- Suppose I am a brain in a vat.
  - (a) Then my assertion of 'I am a brain in a vat' is false.
  - (b) Thus, I am not a brain in a vat – contradiction!
- Therefore, I am not a brain in a vat.

The problem with this argument must be clear by now. The problem is with the derivation of (b) from (a). (a) is correct: my assertion of 'I am a brain in a vat' is false. But that's because I am speaking vat-English in the object language. 'I am a brain in a vat' in vat-English is false iff vat-I is not a vat-brain in a vat-vat. But, indeed, vat-I is not a vat-brain in a vat-vat; it is vat-teaching vat-philosophy, having his vat-brain in his vat-body. So 'I am a brain in a vat' in vat-English is false. However, my metalanguage here is normal-English. Thus, the Disquotation Principle for Falsity does not apply, and (b) is not derivable from (a). End of story.

But what if I am in fact a brain in a vat now and speaking vat-English in the metalanguage as well as in the object language? Then (b) is indeed derivable from (a) by the Disquotation Principle for Falsity. But now the problem is with (a): (a) is not correct and does not follow from the original premise that I am a brain in a vat. For if the metalanguage and the object language are both vat-English, then by the original Disquotation Principle, my assertion of 'I am a brain in a vat' will be true, not false, since I am a brain in a vat. So, again, the argument does not go through. In sum, whether I am speaking normal-English or vat-English in the metalanguage, the argument does not work. I may still be a brain in a vat.

## **PART C: INTERNALIST AND EXTERNALIST ELEMENTS IN PERSONAL IDENTITY**

### **5.14 Locke's Theory of Personal Identity**

We are near the end of the book (yes!). In Part C, I would like to examine John Locke's theory of personal identity and its problems. Locke's theory of

personal identity is a prototypical internalist theory: what takes place in a person's mind determines his/her personal identity. In this part we will see this internalist theory undermined by the need for external elements in the determination of personal identity.

Locke's theory of personal identity is very well known. I am sure most of you have heard about it, if not by that name then in substance. Just as Descartes's skeptical arguments are the basis of many movies about the difference between appearance and reality, Locke's theory is the basis of many novels and movies about amnesia, body change, memory change and transfer, and cloning. They include such classic novels as Stevenson's *Strange Case of Dr Jekyll and Mr Hyde* and Franz Kafka's *Metamorphosis*, and such movies as *Bourne Identity*, *Memento*, *Being John Malkovich*, *Big, Freaky Friday*, *The Shaggy Dog*, *Transcendence*, *The 6th Day*, *Moon*, and *Oblivion*, among many others.

Put simply, Locke's theory of personal identity states that memory makes up personal identity.

- Locke's theory of personal identity (original)  
 $a$  at time  $T1$  is the same person as  $b$  at time  $T2$ , where  $T1 < T2$ , iff  $b$  at  $T2$  remembers  $a$ 's experience at  $T1$ .

Locke's theory of personal identity is the original version of what is usually called the *psychological-continuity theory* or the *memory theory* of personal identity. According to this theory, although Dr Jekyll and Mr Hyde are in the body of one and the same human animal, they are different persons because each does not remember what the other did.

Locke's original definition of personal identity, however, has an obvious flaw. If  $b$  at  $T2$  remembers  $a$ 's experience at  $T1$ ,  $c$  at  $T3$  remembers  $b$ 's experience at  $T2$  but not  $a$ 's experience at  $T1$ , where  $T1 < T2 < T3$ , then  $a$  at  $T1$  is the same person as  $b$  at  $T2$ ,  $b$  at  $T2$  is the same person as  $c$  at  $T3$ , but  $a$  at  $T1$  is not the same person as  $c$  at  $T3$ . For instance (Figure 5.4), Betty at age

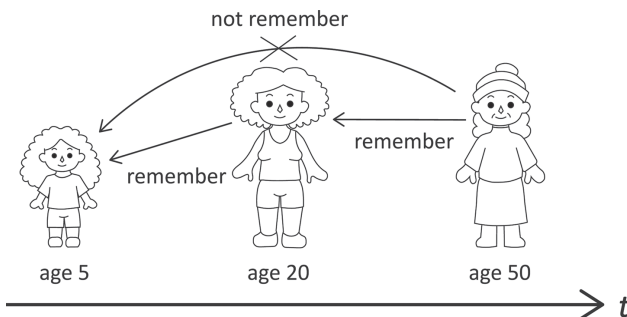


Figure 5.4 Memory link.

20 may remember Betty at age 5, Betty at age 50 may remember Betty at age 20, but Betty at age 50 may no longer remember Betty at age 5.

In that case, Betty at age 20 is the same person as Betty at age 5, Betty at age 50 is the same person as Betty at age 20 but not as Betty at age 5. So there must be two persons overlapping in Betty at age 20.<sup>18</sup> To generalize, Locke's original theory implies that there are numerous persons existing, overlapping one another, in what we usually think of as one person's life.

This is an implication so difficult to accept that Locke's interpreters give Locke a benefit of doubt and interpret his theory as implying that so long as *a* and *b* are connected by a chain of memories, i.e., *b* remembers somebody who remembers somebody who remembers somebody ..., who remembers *a*, *a* and *b* may be considered the same person even if *b* does not remember *a* directly. Let's say that if *b* remembers *a*, there is an ascending memory link from *a* to *b*. Then:

- Locke's theory of personal identity (revised)  
*a* at time *T*<sub>1</sub> is identical with *b* at time *T*<sub>2</sub>, where *T*<sub>1</sub> < *T*<sub>2</sub>, iff there is a chain of ascending memory links from *a* at *T*<sub>1</sub> to *b* at *T*<sub>2</sub>.

This is Locke's theory of personal identity. It is by far the most influential theory of personal identity there is. This theory makes personal identity independent of bodily identity. One person can extend into two or even more bodies so long as they are connected by a single chain of memories. It also makes personal identity independent of the identity of the soul. It really does not matter whether there exist spiritual things called 'souls' separate from physical bodies, or, if there are, whether I keep my soul or lose it at some point; so long as something remembers my experiences, that something is me, and I keep on existing. This makes those movie scenarios possible in which I keep on existing in somebody else's body or even in a machine as a computer program. At the same time, even if I have a soul and it survives the death of my body, if it loses all memory about me, then I don't survive.

This theory is very plausible, for we think we know who we are in the morning in bed even before we open our eyes and look at our bodies; we remember who we are and what we did yesterday or in the past. In some movies, you open your eyes, and you realize that you have a different body from the one you had yesterday. It's still the same you, because you remember who you are. This is an internalist view. What takes place in our minds determines who we are, i.e., our personal identity.

### 5.15 Genuine and Pseudo Memory

However, this internalist theory of John Locke's has been undermined by the need for externalist elements, i.e., elements that are not always accessible

to the subject. I will mention two such elements in this section and the next. Between the two, the first is uncontroversial; the second is based on a more or less controversial response to the so-called *duplication problem*.

The first external element needed is an appropriate causal connection between my memory experience and its origin. Something we all know about memory is that it is often unreliable and can easily be manipulated. Sometimes we seem to remember things that did not happen. I seem to remember that I fell off from the oak tree in the backyard of my parents' house when I was 5 years old just because I have been told about the incident so many times. People's suggestions easily rouse up my fake memory even when I don't really recollect the matter in question. There are people who seriously claim that they are reincarnations of Napoleon Bonaparte or Elvis Presley, and they are happy to tell you stories about what happened to the original Napoleon or Presley. Surely, the fact that they seem to remember Napoleon's/Presley's experiences does not make them Napoleon/Presley.

So we have to make a distinction between *genuine* memory and fake or *pseudo* memory. Locke's definition of personal identity should concern only genuine memories:

- Locke's theory of personal identity (revised and clarified)  
 $a$  at time  $T_1$  is identical with  $b$  at time  $T_2$ , where  $T_1 < T_2$ , iff there is a chain of ascending genuine memory links from  $a$  at  $T_1$  to  $b$  at  $T_2$ .

But how do we distinguish genuine memory from pseudo memory? The most reasonable idea is to define a genuine memory as a memory experience that is appropriately caused by its content. For instance, if my apparent memory of having fallen from a tree is appropriately caused by my actually having fallen from that tree, then my apparent memory is genuine; if not, it is a pseudo memory. Here it is a little difficult to define what 'appropriately caused' really is, but my parents' telling me that I have fallen from the tree is obviously not an appropriate cause (even though that itself might have been caused by my actually having fallen from the tree).

This is somewhat similar to the externalist theory of epistemic justification.

- Genuine memory  
 Subject  $S$  has a genuine memory that  $P$  iff
  - (a)  $S$  has an apparent memory (or memory experience) that  $P$ ;
  - (b)  $P$  is true; and
  - (c)  $S$ 's apparent memory that  $P$  is appropriately caused by the state of affairs that  $P$ .

For  $S$  to have a genuine memory that  $P$ , it is not sufficient for  $S$  to have an apparent memory that  $P$  and for  $P$  to be true; that apparent memory must be appropriately caused by the fact that  $P$ .

However, just like other causal relations, whether there is an appropriate causal relations between my apparent memory and its content is something I usually don't have access to. When I wake up in the morning and 'remember' that I am Ken Akiba, I still can be wrong. So my personal identity is not something I can determine just by looking into my own mind.

### 5.16 The Duplication Problem and the No Competitor Theory

Therefore, Locke's internalist theory of personal identity is unsustainable. In this section – the final section of this book – I would like to mention another way in which the internalist theory of personal identity might break down. This will be a more contentious example than the last because it is based on a more controversial theory.

One problem Locke's theory of personal identity has, aside from that about the distinction between genuine memory and pseudo memory, is the so-called *duplication problem*. Think about the following sci-fi scenario (Scenario 1): My memory in my normal brain is transplanted by surgery into another, newly created brain inside a cloned body of mine, while my original body is destroyed. After this surgery, I wake up in the cloned body. No problem; the person in the cloned body is me, genuinely remembering Ken Akiba's past experiences. Think, however, of the following situation next (Scenario 2): instead of my memory's being transplanted into one brain in one cloned body, it is transplanted into two brains in two cloned bodies. Suppose that the two persons come out of the surgery not knowing that there is another clone. So each of them thinks that he is Ken Akiba. Then how should we describe the situation? Between these two persons coming out of the surgery, which is Ken Akiba? Perhaps both? Perhaps neither?

Let's call the person going into the surgery in both scenarios, *a*, the person coming out of the surgery in Scenario 1, *b*, the two persons coming out of the surgery in Scenario 2, *b1* and *b2*. One thing we want to make true is that in Scenario 1, *a* and *b* are one and the same person (i.e., *a* and *b* are parts of one and the same four-dimensional person). Another thing we should keep in mind is that in Scenario 2, everything is symmetric between *b1* and *b2*; so if *b1* is the same person as *a*, so is *b2*.

Given these constraints, there are at least three theoretical options about Scenario 2 while *a* is considered the same person as *b* in Scenario 1:

- (1) *a*, *b1*, and *b2* are all one and the same person;
- (2) *a* and *b1* are one and the same person, *a* and *b2* are one and the same person, but *b1* and *b2* are not one and the same person; that is, *a* is a part of two persons;
- (3) *a*, *b1*, and *b2* are all different persons.

However, many people find option (1) weird and totally unacceptable. If  $b_1$  and  $b_2$  are one and the same person, and if they talk to each other after the surgery, is the person in question talking to himself? Also,  $b_1$  and  $b_2$  are going to go on their own ways, live different lives, and have different experiences after the surgery. How can we still consider them one and the same person?

Many people also find option (2) weird. It means that there are two persons already in  $a$ 's body, sharing the same thoughts, memories, and stream of consciousness; and  $a$  does not even know it.

If you don't want to accept either (1) or (2), the only option left is (3), according to which person  $a$  is killed at the point of the surgery, and the new persons  $b_1$  and  $b_2$  are born. But how is that possible? After all,  $b_1$  is just like  $b$  had it not been for the existence of  $b_2$ , and vice versa. *The no competitor theory* maintains, however, that the other person  $b_2$ , a competitor, works as a defeater of the personal identity between  $a$  and  $b_1$ . When we say that  $b_1$  and  $b_2$  are competitors for the same person as  $a$ , we mean that  $b_1$  and  $b_2$  have equally strong claims to be the same person as  $a$ . The competitor theory says that a competitor should be a defeater because, otherwise, we have to accept one of the untenable options (1) and (2). So:

- The No Competitor Theory  
 $a$  at  $T_1$  is the same person as  $b$  at  $T_2$ , where  $T_1 < T_2$ , iff
  - (a) there is a chain of ascending genuine memory links from  $a$  at  $T_1$  to  $b$  at  $T_2$ ; and
  - (b) neither  $a$  at  $T_1$  nor  $b$  at  $T_2$  has a competitor.<sup>19</sup>

I find the no competitor theory a rather weak theory because there is no good reason why the competitors should be defeaters of personal identity, except that, otherwise, you'd have to accept one of the more distasteful options (1) and (2). There is no *inherent* motivation, or deeper reason, to embrace the no competitor theory. However, the reason why I mentioned this theory here is because it introduces another externalist element to the determination of personal identity. If the theory is correct, then in order for me to know that I am Ken Akiba, it is not sufficient for me to have the apparent memory that I am Ken Akiba; it is not sufficient for me to know, furthermore, that that apparent memory is appropriately caused. I have to know that there is nobody else who has an equally strong claim to be Ken Akiba. If I alone come out of the surgery, I am Ken Akiba. But if a second person is later created, then I will turn out not to be Ken Akiba.

I will leave it to the reader to decide whether this option is really better than option (2), or even option (1). Note that even option (2) implies that I do not always know how many persons are in my body, and option (1) implies

that I do not always know whether the person I am interacting with is myself. So, regardless of which of the three options is the correct answer, the duplication problem shows that there are things about our personal identity that we cannot know from inside.

I should add that the no competitor theory is applicable to many instances of temporal identity (or identity over time), not just personal identity, which is a specific instance of temporal identity. Think of the following variation of the Ship of Theseus puzzle, discussed earlier in Section 1.5. In the TV series *Star Trek*, people and other objects are transported (or ‘beamed down’) from the starship USS Enterprise to the surfaces of stars and planets. A crew pulls the lever, and the object placed in the transport station disappears, and apparently the same object will appear down on the ground. If one object, say a small spaceship, is transported from Enterprise to the ground, we will consider it the same spaceship. But what if the transporter malfunctions and two identical (i.e., indistinguishable) spaceships appear down on the ground? Which ship is the same ship as the original?

If the no competitor theory is correct, then neither is the same ship as the original, while if only one ship appears, that’s the same ship as the original. Again, it is debatable whether this conclusion is better than the other options; but it is clear that there is a close analogy between this case and the personal identity case.

To summarize the last two sections: Locke’s theory of personal identity initially seemed plausible because it is supported by our strong feeling that we know who we are from inside. The last two sections have shown, however, that this view cannot be sustained as is, and that certain externalist elements must be at least added to the internalist elements to give a fully satisfactory account of personal identity. It is an interesting question to ask whether a purely externalist theory of personal identity is possible or plausible.

Animalism (Olson 1997) is a theory of personal identity that rivals the psychological-continuity theory. According to animalism, a human person is a stage of a human animal during which s/he retains self-consciousness and rationality. While this condition does not exclude the possibility of non-human persons, it excludes the possibility of one human person’s persisting outside the body of one human animal; that is, we don’t survive through cloning or memory transfer. Consequently, *a*, *b1*, and *b2* in the duplication case are all different persons; even *a* and *b* are not the same person. On the other hand, Jekyll and Hyde are one and the same person because they are the same animal. However, animalism, interesting as it may be, is still a fledgling minority view compared to the psychological-continuity theory, and its plausibility has yet to be determined.



### Exercise Questions

1. Explain the following concepts and distinctions.  
Externalism/internalism; semantic internalism/externalism; the description theory/the causal theory; narrow/wide content; intrinsic/extrinsic properties; intrinsic/extrinsic value; supervenience; JTB theory of knowledge; Gettier problem; internalism/externalism about epistemic justification; foundationalism/coherentism/reliabilism; Disquotation Principle(s); psychological-continuity (or memory) theory of personal identity; genuine/pseudo memory; duplication problem; no competitor theory.
2. Describe Kripke's metaphysical, semantic, and epistemological arguments against the description theory.
3. Describe Searle's replies to Kripke's metaphysical and semantic arguments.
- \*4. Which of the following are natural kinds?  
(a) water (b) beverage (c) mammal (d) flying animal (e) lion (f) panther (g) rose (h) weed (i) rice (j) bread (k) mountain (l) friend (m) art
- \*5. Which of the following are intrinsic properties, and which are extrinsic properties? Some contentious examples may be included.  
(a) being the First Lady (b) being a dog (c) having a girlfriend (d) being blonde (e) being famous (f) being 6 feet tall (g) being tall (h) having a high IQ
6. Describe Putnam's externalist argument against the 'brain in a vat' hypothesis, and explain how it does not work.

### Discussion Questions

1. In his criticism of the description theory, Kripke assumes that even when we don't have an appropriate (non-parasitic) definite description, we successfully refer to famous historical figures like Einstein and Aristotle by their names. Is this assumption plausible? Is there any case in which our misunderstanding or lack of understanding of the owner of the name is so severe that we cannot be considered to be successfully referring to the owner?
2. There are very common personal proper names such as 'John', 'Elizabeth', 'Maria', and 'Smith'. You may know more than one person named 'John' or 'Elizabeth'. When you say 'John' and refer to one John rather than another, what is happening? What would

- be the descriptivist's account of the situation? What would be the causal theorist's account of the situation? Which theory gives a better account?
3. Suppose your intuition says that Oscar and T'Oscar have different mental contents while your friend's intuition says that they have the same mental content. How should we determine whose intuition is correct? How should we, or should we not, settle the issue? How important is this sort of intuition in philosophical discussions?
  4. What do you think of the solution, suggested in Section 5.9, to the problem of grounding raised in Section 3.8? Are you satisfied with the solution? Why or why not?
  5. Why do you think that Earth is round (or spherical)? Formulate the justificational structure of beliefs that support your belief that Earth is round, i.e., what beliefs and non-doxastic states support your belief that Earth is round, and what beliefs and non-doxastic mental states support those beliefs, etc. Does the resulting structure conform more to foundationalism or coherentism?
  6. I said in Section 5.12 that my belief that Justify has won was justified because the belief was produced by a reliable process, perception, while my earlier belief that Justify would win was not justified because it was produced by an unreliable process, pure guessing. Can you give an internalist (foundationalist or coherentist) account of why the former belief is justified while the latter belief is unjustified? Is the externalist (reliabilist) account really better than the internalist account?
  7. What virtues do beliefs have which are internally justified? What virtues do beliefs have which are externally reliable? Do they have different virtues?
  8. What do you think about the duplication problem in personal identity? Is it a problem only for internalist theories of personal identity, or does it also apply to externalist theories? What do you think about the no competitor solution? Do you find it plausible? Why or why not? Do you agree that option (3) in the text is better than option (1) or (2)?

## Suggested Further Reading

The following are useful encyclopedia articles on the title topics:

- Joe Lau and Max Deutsch, "Externalism about Mental Content."
- Dan Marshall and Brian Weatherson, "Intrinsic vs. Extrinsic Properties."
- Brian McLaughlin and Karen Bennett, "Supervenience."

They may seem too detailed at places on the first reading, but should give you good ideas about the entire fields.

Searle's response to Kripke's criticism of the description theory of proper names is included in

- John Searle, *Intentionality*, Chapter 9: Proper Names and Intentionality (reprinted in A. P. Martinich and David Sosa (eds.), *The Philosophy of Language*, 6th edn).

If you are looking for a standard textbook in epistemology,

- Matthias Steup, *An Introduction to Contemporary Epistemology*

may be recommended.

- John Perry (ed.), *Personal Identity*, 2nd edn

is a new edition of an excellent anthology on personal identity.

## Notes

- 1 There are some exceptions, however. For instance, *motivational internalism* in moral psychology states that our conviction that  $x$  (morally) ought to be done is essentially motivating for us to do  $x$ , while *motivational externalism* denies this, allowing the possibility of an amoralist who sincerely believes that  $x$  ought to be done, yet is him/herself not at all motivated to do  $x$ . The former position holds that our moral beliefs and motivations have an internal connection while the latter holds that the relation is only external. Here the distinction 'internal/external' is not used as 'internal/external to the mind' (even though the whole topic has much to do with mind).
- 2 A perceptive reader may notice that the description 'the teacher of Alexander the Great' itself contains a proper name 'Alexander the Great'. So we have to replace this name with another description. Does this get into an infinite regress or a vicious circle? Russell himself thought that this regress will end at one point, where what he calls 'logically proper names', which are not abbreviations of definite descriptions, denote things we are directly acquainted with (by perception, etc.). For the possibility of vicious circularity, see Kripke's criticisms of the description theory in the next section.
- 3 Kurt Gödel is a real logician, one of the greatest logicians of the twentieth century, who in fact proved the incompleteness of arithmetic. Kripke and Gödel were colleagues at Princeton University when Kripke gave a lecture series on which his *Naming and Necessity* is based.
- 4 Actually, it has been reported (Machery, Mallon, Nichols, and Stich 2004) that East Asians are less likely than Westerners to share this intuition.

- 5 Searle also makes a good point in saying that the internalist about the reference of proper names does not have to be a descriptivist, who holds the view that the reference of a proper name is determined solely by the definite description(s) the speaker associates with the name. Some other internalist elements than descriptions may be available to the internalist, such as the speaker's perception of the denotation, as the determinants of the name's denotation. Perception may play a larger role than descriptions when we deal with the proper names of ordinary people like you and me, as opposed to those of famous historical figures like Einstein and Aristotle. I may not have a definite description of you handily available which can single you out from all the other students (sorry!), but I can tell it's you when I see you.

Searle still retains his cluster concept theory but incorporates these elements (parasitic descriptions, perception, etc.) into the relevant cluster.

- 6 Another influential philosopher who made the same point about rigidification is Plantinga (1974). Davies and Humberstone (1980) introduced the 'actually' operator in modal logic and also made the same point.

There is also a uniform way to construct a rigid definite description from a proper name '*N*': just add 'the thing identical with' in front of '*N*', creating the definite description 'the thing identical with *N*'. So if '*N*' is 'Aristotle', the definite description 'the thing identical with Aristotle' is a rigid designator, designating Aristotle in all possible worlds in which Aristotle exists. This may seem a little like cheating, but there is no doubt that the resulting description is a rigid designator.

- 7 There are kinks in Putnam's scenario that need to be fixed. Water is prevalent on Earth and constitutes about 60 percent of Oscar's body; similarly for t'water, Twin Earth, and T'Oscar. So it is just not possible that Earth and Twin Earth have parallel history until 1750 and that Oscar and T'Oscar have the same mental state at that point. The difference between H<sub>2</sub>O and XYZ should show up somewhere. So a more rare substance like gold should make a better example, even though we will continue using water in our example.
- 8 One of my students gave me a similar but actual example: While tiger is a natural kind, panther is not; panthers in Asia and Africa are (black) leopards but those in the Americas are jaguars, cougars, or bobcats.
- 9 Here I identify *dispositional* properties as modal properties; for instance, the dispositional property *x* is *soluble* may be identified as the modal property *x* is *dissolved if put in water*.
- 10 Another kind often mentioned is *knowledge by acquaintance*. If I say 'I know Adam', what I mean is that I *am acquainted with* him. Many philosophers say that knowledge of this kind is not reducible to know-how or know-that.
- 11 Justify won the Triple Crown (i.e., Kentucky Derby, Preakness Stakes, and Belmont Stakes) in 2018.
- 12 The internalism which makes this claim is sometimes called *access internalism* and is distinguished from *ontological internalism*, which holds only that whether a belief is justified or not is determined solely by the subject's mental states. Ontological internalism does not commit itself to the further claim that the subject has access to those mental states on reflection. We will focus on access internalism in this book.

13 A more careful definition goes as follows:

*S*'s belief *B* is justified iff

- (a) *B* is produced by a reliable cognitive process; and
- (b) there is no other reliable cognitive process available to *S* which, had it been used by *S* in addition to the process actually used, would have resulted in *S*'s not forming *B*.

For the reason why clause (b) should be added, see the next note. But for an initial understanding of the theory, the simpler version presented in the main text will do.

14 However, there is a little complication here. If Norman at the same time watches live on TV the President's landing in a foreign country, then his belief that the President is at the White House is not justified and should be withdrawn, assuming that TV watching is a more reliable cognitive process than clairvoyance. One reliable cognitive process may be superseded by another, even more reliable, cognitive process. That's why clause (b) is needed in the more careful definition of justification given in the last note.

The clairvoyance example was originally presented by Bonjour (1985) as a criticism of reliabilism. I use it here not as a criticism but as a statement of the fact that reliabilism, as externalism, has this consequence. The reader should decide on his/her own if this consequence is untenable.

Even though the clairvoyance example is (at this point) a fiction, we may well have actual similar cases in our hands. For instance, according to the most recent research, not just many non-human animals (e.g., migratory birds and sea turtles) but even some humans have magnetoreception, the sense that detects Earth's magnetic field. If one of those humans, who is in a group of hikers lost in a forest, points to one direction and says 'That's north', she herself may take this as a mere hunch, but she may be justified in believing so in the reliabilist sense of justification.

15 Or, more generally:

If *S* knows that *P*, then *S* knows that *S* knows that *P*.

16 Even though, even in that case, there is an apparent gap between possibility and actuality: the justification's being accessible to me does not mean that I always make contact with it and come to believe that I am justified.

17 In Section 2.1, note 1, I mentioned the disquotational (or deflationary) theory of truth. Put simply, the disquotational theory maintains that the role of the truth predicate is exhausted by what's stated in the Disquotation Principle, and that it is a mistake to think that there is some substantive property, truth, in the world denoted by the predicate. Since representational (mental and linguistic) content is characterized as the kind of thing which can have this substantive property of truth, the disquotational theory denies the existence of representational content and draws a radically different picture of language and mind than the correspondence theory of truth, on which this entire book is based.

18 The Scottish philosopher Thomas Reid (1785) and many others following him argued that Locke's theory was in fact inconsistent, for if *a* is the same person as *b* and *b* is the same person as *c*, then by the transitivity of identity, *a* must be

the same person as *c*. That does not follow, however, because the predicate '*x* is the same person as *y*' need not be the predicate of numerical identity (Section 1.5). If it were the predicate of numerical identity, then, of course, Betty at age 20 would not be the same person as Betty at age 5 (or 50) because the temporal part Betty-at-age-20 (Section 3.16) is a numerically different object from the temporal part Betty-at-age-5 (or 50). Among other things, they exist at different time points. Instead, the predicate '*x* is the same person as *y*' in question may be understood as something like 'there is at least one four-dimensional person that *x* and *y* are both parts of'. But then there is no inconsistency in the claim that Betty at age 20 is the same person as Betty at age 5, that Betty at age 20 is the same person as Betty at age 50, but that Betty at age 5 is not the same person as Betty at age 50. Transitivity does not generally hold for this sense of identity.

- 19 For instance, Nozick's (1981) *closest continuer theory* of personal identity is a version of the no competitor theory.

---

# Answers to Selected Exercise Questions

---

## Chapter 1

3. (a)  $L \wedge H$   
(b)  $L \vee S$   
(c)  $H \leftrightarrow L$   
(d)  $\neg H \rightarrow S$   
(e) If I am sad, then my dog doesn't love me (or, I am sad only if my dog doesn't love me). False.  
(f) If I am happy or I am sad, my dog loves me. True.  
(g) If I am happy if my dog loves me, then either my dog loves me or I am sad. True.  
(h) If I am happy and sad, then I am happy and not happy. False.

4. (a)

$P$	$\neg$	$\neg$	$P$
T	<b>T</b>	F	T
F	<b>F</b>	T	F

- (b)

$P$	$P$	$\leftrightarrow$	$\neg$	$\neg$	$P$
T	T	<b>T</b>	T	F	T
F	F	<b>T</b>	F	T	F

- (c)

$P$	$P$	$\wedge$	$\neg$	$P$
T	T	<b>F</b>	F	T
F	f	<b>F</b>	T	F

(d)

P	P	$\vee$	$\neg$	P
T	T	<b>T</b>	F	T
F	F	<b>T</b>	T	F

(e)

P	Q	$\neg$	P	$\wedge$	Q
T	T	F	T	<b>F</b>	T
T	F	F	T	<b>F</b>	F
F	T	T	F	<b>T</b>	T
F	F	T	F	<b>F</b>	F

(f)

P	Q	$\neg$	(P	$\wedge$	Q)
T	T	<b>F</b>	T	T	T
T	F	<b>T</b>	T	F	F
F	T	<b>T</b>	F	F	T
F	F	<b>T</b>	F	F	F

(g)

P	Q	(P	$\wedge$	Q)	$\leftrightarrow$	(P	$\rightarrow$	$\neg$	Q)
T	T	T	T	T	<b>F</b>	T	F	F	T
T	F	T	F	F	<b>F</b>	T	T	T	F
F	T	F	F	T	<b>F</b>	F	T	F	T
F	F	F	F	F	<b>F</b>	F	T	T	F

5. (a)  $Ta \wedge \neg Tb$   
 (b)  $Tb \rightarrow \neg Sba$   
 (c)  $Ta \vee (Tb \vee Tc)$  (or  $(Ta \vee Tb) \vee Tc$ , or  $Ta \vee Tb \vee Tc$ )  
 (d)  $\neg Sca \wedge \neg Scb$   
 (e)  $Sca \rightarrow Scb$   
 (f) Buster is shorter than Chris, or Chris is shorter than Buster. True.  
 (g) If Chris is tall, then Buster is shorter than Chris. True.  
 (h) It is not the case that Adam and Chris are both shorter than Buster. True.  
 (i) Adam or Buster is tall, and Buster or Chris is tall. False.  
 (j) If Buster is tall or Chris is shorter than Buster, Buster is tall. False.

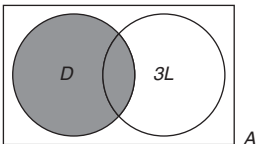


6. (a) (A) Necessary condition (B) Neither  
 (b) (A) Sufficient condition (B) Sufficient condition  
 (c) (A) Necessary condition (B) Necessary and sufficient condition
7. (a) The most famous disciple of Plato is Aristotle. Unnecessary.  
 (b) The most famous disciple of Plato is identical with Aristotle. Unnecessary.  
 (c) "The most famous disciple of Plato" refers to Aristotle.  
 (d) "The most famous disciple of Plato" refers to the same thing as "Aristotle" does.  
 (e) The most famous disciple of Plato and Aristotle are one and the same man. Unnecessary.  
 (f) The denotation of "the most famous disciple of Plato" is Aristotle.  
 (g) John said, "The most famous disciple of Plato is Aristotle."  
 (h) John said, "'The most famous disciple of Plato' denotes Aristotle."  
 (i) John told Ann that "the most famous disciple of Plato" denoted Aristotle.  
 (j) Being tall and being a dog are both properties. Unnecessary.  
 (k) "Is tall" and "is a dog" are both predicates.  
 (l) "The most famous disciple of Plato is Aristotle" is true.  
 (m) It is true that the most famous disciple of Plato is Aristotle. Unnecessary.  
 (n) The post office has thousands of letters sent every day. Unnecessary.  
 (o) "The post office" has only 12 letters.  
 (p) Adam knows Karate. Unnecessary.  
 (q) Adam knows "Karate" and three other Japanese words.
8. The song, whose lyrics are (or begin with) '*A-sitting On A Gate*', has the official name, '*The Aged Aged Man*'. But the name itself has a name, '*Haddocks' Eyes*'. The song also has an unofficial name (or is commonly called) '*Ways and Means*'.

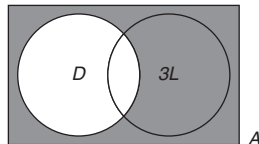
## Chapter 2

3.

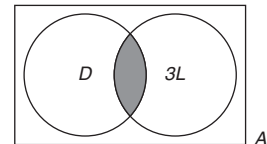
(a)



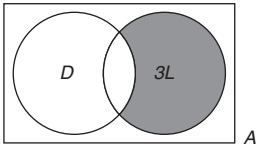
(b)



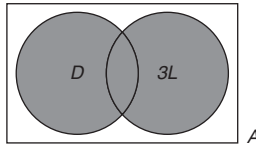
(c)



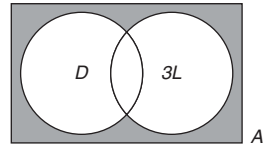
(d)



(e)



(f)



- (g) The set of female dogs that are not three-legged.  
 (h) The set of dogs and female animals either of which are not three-legged.  
 (i) The set of male animals except dogs.  
 (j) The set of male three-legged dogs and female animals that are neither dogs nor three-legged.
6. (a) The '-1' function (i.e.,  $y = x-1$ ). (This is just one example; there are other functions which give the same values for the same arguments. Same below.)  
 (b) The cube function (i.e.,  $y = x^3$ ).  
 (c)  $y = -2x$ .  
 (d)  $y$  is the most famous disciple of  $x$ .  
 (e)  $y$  is the President of the US immediately preceding  $x$ .  
 (f)  $y$  is the truth value of (e.g.) ' $x$  is a philosopher'.

### Chapter 3

3. Logical truths: (b) and (d). Logical falsities: (c) and (g).
5. If  $Q$  is a logical truth, then it is impossible for  $Q$  to be false; so, for any  $P$ , it is impossible for  $P$  to be true and  $Q$  false; therefore, the argument is valid. If  $P$  is a logical falsity, then it is impossible for  $P$  to be true; so, for any  $Q$ , it is impossible for  $P$  to be true and  $Q$  false; therefore, the argument is valid. QED.
6. (a) Logical; analytic.  
 (b) Logical; analytic.  
 (c) Non-logical; analytic.  
 (d) Non-logical; synthetic.  
 (e) Non-logical; analytic.  
 (f) Logical; analytic.  
 (g) Logical; analytic.  
 (h) Logical; analytic.  
 (i) Logical; analytic.  
 (j) Non-logical; synthetic.

9. (a) If it has always been the case that  $P \rightarrow Q$ , then if it has always been the case that  $P$ , it has always been the case that  $Q$ . Logically true.
- (b) If it is going to be always the case that  $P \rightarrow Q$ , then if at some time in the future,  $P$ , then at some time in the future,  $Q$ . Logically true.
- (c) At no time in the past,  $P$ , iff it has always been the case that  $\neg P$ . Logically true.
- (d) It is not going to be always the case that  $P$  iff at some time in the future,  $\neg P$ . Logically true.
- (e) If at some time in the past,  $P$ , and at some time in the past,  $Q$ , then at some time in the past,  $P \wedge Q$ . Not logically true.
- (f) If at some time in the future,  $P$ , or at some time in the future,  $Q$ , then at some time in the future,  $P \vee Q$ . Logically true.
- (g) If  $P$  now, then it is going to be always the case that at some time in the past,  $P$ . Logically true.
- (h) If it has always been the case that at some time in the future,  $P$ , then either  $P$  now or  $P$  at some time in the future. Logically true.
- (i) If it is going to be always the case that  $P$ , then at some time in the future,  $P$ . Logically true (assuming that by its nature, time continues to exist after now).

## Chapter 4

2. (a) The world is flat.
- (b) Non-representational.
- (c) Someone loves Adam.
- (d) You came.
- (e) Non-representational. (The answers to (d) and (e) are debatable.)
- (f) The truck runs over the dog.
- (g) Charlie is rich.
- (h) Non-representational. (Most philosophers probably will say that (b) and (h) are non-representational.)
- (i) You will come to the party.
3. (a) No (b) Yes (c) Yes (d) Yes (at least in some circumstances)
- (e) No (f) Yes (g) No (h) Yes? (It is debatable whether the referees in sports games create facts, e.g., about whether the player is safe or out, the ball is a strike or ball, in or out, etc., by declaration, or they try to describe facts that exist independently and objectively. I think that, thanks to many recent technological advances, more and more people think of it in the latter way these days.)

**Chapter 5**

4. (a), (c), (e), (g), and (i). (Re (f), see note 8.)
5. (a) Extrinsic. (b) Intrinsic. (c) Extrinsic. (d) Intrinsic. (e) Extrinsic. (f) Intrinsic. (g) Extrinsic (assuming that being tall implies being taller than the average). (h) This is a little complicated; having a certain IQ is intrinsic, but that IQ's being high is extrinsic, relative to the others' IQs.

---

# Bibliography

---

- Adams, R. M. (1974) Theories of actuality. *Noûs* 8: 211–231.
- Anscombe, G. E. M. (1956) Aristotle and the sea battle. *Mind* 65: 1–15.
- Armstrong, D. M. (1978) *Universals and Scientific Realism*, 2 vols. Cambridge, England: Cambridge University Press.
- Armstrong, D. M. (1989) *Universals*. Boulder, CO: Westview.
- Austin, J. L. (1962) *How to Do Things with Words*, 2nd edn. Oxford: Clarendon Press.
- Ayer, A. J. (1936) *Language, Truth, and Logic*. London: Gollancz.
- Ayer, A. J. (1940) *The Foundations of Empirical Knowledge*. London: Macmillan.
- Barwise, J., and Perry, J. (1983) *Situations and Attitudes*. Cambridge, MA: MIT Press.
- Berkeley, G. (1713) *Three Dialogues between Hylas and Philonous in Opposition to Sceptics and Atheists*. London: Printed by G. James, for Henry Clements.
- Black, M. (1952) The identity of indiscernibles. *Mind* 61: 153–164.
- Boghossian, P. A. (1997) Analyticity. In B. Hale and C. Wright (eds.), *A Companion to the Philosophy of Language*, 331–368. Oxford: Blackwell.
- BonJour, L. (1985) *The Structure of Empirical Knowledge*. Cambridge, MA: Harvard University Press.
- Broadbent, A. (2016) *Philosophy for Graduate Students*. New York: Routledge.
- Burge, T. (1979) Individualism and the mental. *Midwest Studies in Philosophy* 4: 73–121.
- Carnap, R. (1947) *Meaning and Necessity*. Chicago: University of Chicago Press.
- Carroll, L. (1871) *Through the Looking-Glass, and What Alice Found There*. London: Macmillan.
- Casati, R., and Varzi, A. C. (1999) *Parts and Places*. Cambridge, MA: MIT Press.
- Chalmers, D. J. (1996) *The Conscious Mind*. Oxford: Oxford University Press.
- Chalmers, D. J. (2004) Epistemic two-dimensional semantics. *Philosophical Studies* 118: 153–226.
- Chalmers, D. J. (2006) Two-dimensional semantics. In E. Lepore and B. Smith (eds.), *The Oxford Handbook of Philosophy of Language*, 574–606. Oxford: Clarendon Press.
- Chalmers, D. J. (2010) *The Character of Consciousness*. Oxford: Oxford University Press.
- Chisholm, R. (1942) The problem of the speckled hen. *Mind* 51: 368–373.
- Churchland, P. M. (1985) Reduction, qualia and direct introspection of brain states. *Journal of Philosophy* 82: 8–28.

- Cresswell, M. J. (1985) *Structured Meanings*. Cambridge, MA: MIT Press.
- Davidson, D. (1968) On saying that. *Synthese* 19: 130–146.
- Davidson, D. (1970) Mental events. In L. Foster and J. W. Swanson (eds.), *Experience and Theory*, 79–101. London: Duckworth.
- Davies, M., and Humberstone, I. L. (1980) Two notions of necessity. *Philosophical Studies* 58: 1–30.
- Descartes, R. (1637) *Discourse on Method / Discours de la méthode*. Leiden: Ian Maire.
- Descartes, R. (1641) *Meditations on First Philosophy / Meditationes de prima philosophia*. Paris: Michael Soly
- Donnellan, K. S. (1970) Proper names and identifying descriptions. *Synthese* 21: 335–358.
- Evans, G. (1979) Reference and contingency. *Monist* 62: 161–189.
- Fodor, J. (1975) *The Language of Thought*. Cambridge, MA: Harvard University Press.
- Frege, G. (1892) Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50.
- Geach, P. T. (1965) Assertion. *Philosophical Review* 74: 449–465.
- Gettier, E. (1963) Is justified true belief knowledge? *Analysis* 23: 121–123.
- Goble, L. (ed.) (2001) *The Blackwell Guide to Philosophical Logic*. Malden, MA, and Oxford: Blackwell.
- Goldman, A. (1979) What is justified belief? In G. S. Pappas (ed.), *Justification and Knowledge*, 1–23. Dordrecht: Reidel.
- Hare, R. M. (1952) *The Language of Morals*. Oxford: Oxford University Press.
- Hawthorne, J. (2003) Identity. In M. J. Loux and D. W. Zimmerman (eds.), *The Oxford Handbook of Metaphysics*, 99–130. Oxford: Oxford University Press.
- Hume, D. (1748) *An Enquiry Concerning Human Understanding*. London: Printed for A. Millar.
- Jackson, F. (1982) Epiphenomenal qualia. *Philosophical Quarterly* 32: 127–136.
- Jackson, F. (1998) *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- Kant, I. (1783) *Prolegomena to Any Future Metaphysics / Prolegomena zu einer jeden künftigen Metaphysik*. Riga: Johann Friedrich Hartknoch.
- Kleene, S. C. (1952) *Introduction to Metamathematics*. Amsterdam: North-Holland.
- Kripke, S. (1980) *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Kroon, F. W. (1987) Causal descriptivism. *Australasian Journal of Philosophy* 65: 1–17.
- LaPorte, J. (2016) Rigid designators. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/rigid-designators>.
- Lau, J., and Deutsch, M. (2014) Externalism about mental content. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/content-externalism>.
- Lehrer, K. (1974) *Knowledge*. Oxford: Oxford University Press.
- Lewis, D. (1973a) *Counterfactuals*. Oxford: Blackwell.
- Lewis, D. (1973b) Causation. *Journal of Philosophy* 70: 556–567.
- Lewis, D. (1986) *On the Plurality of Worlds*. Oxford: Blackwell.
- Locke, J. (1690) *An Essay Concerning Human Understanding*. London: Printed by Eliz. Holt, for Thomas Basset.
- Loux, M. J. (ed.) (1979) *The Possible and the Actual*. Ithaca, NY: Cornell University Press.

- Machery, E., Mallon, R., Nichols, S., and Stich, S. P. (2004) Semantics, cross-cultural style. *Cognition* 92: B1–B12.
- Mackie, J. L. (1977) *Ethics*. Harmondsworth: Penguin.
- Marshall, D., and Weatherson, B. (2018) Intrinsic vs. extrinsic properties. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/intrinsic-extrinsic>.
- Martinich, A. P., and Sosa, D. (eds.) (2013) *The Philosophy of Language*, 6th edn. Oxford: Oxford University Press.
- McLaughlin, B., and Bennett, K. (2018) Supervenience. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/supervenience>.
- Mellor, D. H., and Oliver, A. (eds.) (1997) *Properties*. Oxford: Oxford University Press.
- Mill, J. S. (1843) *A System of Logic*, 2 vols. London: John W. Parker.
- Musgrave, A. (1993) *Common Sense, Science and Scepticism*. Cambridge, England: Cambridge University Press.
- Nagel, T. (1974) What is it like to be a bat? *Philosophical Review* 83: 435–450.
- Nelson, M. (2019) Propositional attitude reports. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/prop-attitude-reports>.
- Ney, A. (2014) *Metaphysics*. London: Routledge.
- Nolan, D. (2013) Impossible worlds. *Philosophy Compass* 8: 360–372.
- Nolt, J., Rohatyn, D., and Varzi, A. (2011) *Logic*, 2nd edn. New York: McGraw-Hill.
- Nozick, R. (1981) *Philosophical Explanations*. Cambridge, MA: Harvard University Press.
- Olson, E. T. (1997) *The Human Animal*. New York: Oxford University Press.
- Perry, J. (ed.) (2008) *Personal Identity*, 2nd edn. Berkeley: University of California Press.
- Plantinga, A. (1972) *The Nature of Necessity*. Oxford: Clarendon Press.
- Priest, G. (2008) *An Introduction to Non-Classical Logic*, 2nd edn. Cambridge, England: Cambridge University Press.
- Putnam, H. (1975) The meaning of ‘meaning’. *Minnesota Studies in the Philosophy of Science* 7: 131–193.
- Putnam, H. (1981) *Reason, Truth and History*. Cambridge, England: Cambridge University Press.
- Quine, W. V. O. (1951) Two dogmas of empiricism. *Philosophical Review* 60: 20–43.
- Quine, W. V. O. (1956) Quantifiers and propositional attitudes. *Journal of Philosophy* 53: 177–187.
- Ravenscroft, I. (2005) *Philosophy of Mind: A Beginner's Guide*. Oxford: Oxford University Press.
- Rea, M. (ed.) (1997) *Material Constitution*. Lanham, MD: Rowman & Littlefield.
- Reid, T. (1785) *Essays on the Intellectual Powers of Man*. Dublin: Printed for L. White.
- Russell, Bertrand (1905) On denoting. *Mind* 14: 479–493.
- Russell, Bertrand (1912) *The Problems of Philosophy*. London: Williams & Norgate.
- Russell, Bertrand (1919) *Introduction to Mathematical Philosophy*. London: George Allen & Unwin.
- Russell, Bruce (2014) A priori justification and knowledge. In E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, <https://plato.stanford.edu/entries/apriori>.



- Searle, J. R. (1958) Proper names. *Mind* 67: 166–173.
- Searle, J. R. (1983) *Intentionality*. Cambridge, England: Cambridge University Press.
- Sellars, W. (1956) *Empiricism and the Philosophy of Mind*. Cambridge, MA: Harvard University Press.
- Sosa, E., and Tooley, M. (eds.) (1993) *Causation*. Oxford: Oxford University Press.
- Stalnaker, R. (1968) A theory of conditionals. In N. Rescher (ed.), *Studies in Logical Theory*, 98–112. Oxford: Oxford University Press.
- Stalnaker, R. (1976) Propositions. In A. F. Mackey and D. D. Merrill (eds.), *Issues in the Philosophy of Language*, 79–91. New Haven, CT: Yale University Press.
- Stalnaker, R. (1984) *Inquiry*. Cambridge, MA: MIT Press.
- Steup, M. (1996) *An Introduction to Contemporary Epistemology*. Upper Saddle River, NJ: Prentice-Hall.
- Stevenson, C. L. (1937) The emotive meaning of ethical terms. *Mind* 46: 14–31.
- Suppes, P. (1972) *Axiomatic Set Theory*. New York: Dover.
- van Fraassen, B. C. (1968) Presupposition, implication and self-reference. *Journal of Philosophy* 65: 136–152.
- Williams, D. C. (1953a) On the elements of being: I. *Review of Metaphysics* 7: 3–18.
- Williams, D. C. (1953b) On the elements of being: II. *Review of Metaphysics* 7: 171–192.



---

# Index

---

- aboutness 106  
abstract object 2–5, 30n1, 34, 35,  
41, 42, 62  
Adams, R. M. 63n10, 177  
adjective 3, 10, 24, 42  
adverb 10, 22, 122  
analyticity 64, 78, 99, 177  
analytic philosophy 21  
animalism 164  
Anscombe, G. E. M. 24, 177  
antecedent 6, 88, 103, 104, 123, 155  
anti-realism 4, 5, 26, 116; mathematical 5;  
modal 48; moral 5, 118, 121, 144  
a posteriori 44, 68, 69, 70, 78, 82, 98, 99,  
103, 134, 138; necessary 100, 101n8;  
*see also* knowledge; proposition; truth  
a priori 44, 65, 68, 69, 70, 78, 81, 84,  
98, 99, 103, 138, 179; contingent  
99, 101n8; synthetic 82, 100; *see also*  
knowledge; proposition; truth  
argument 5, 6, 32, 44, 49, 50, 56, 57, 60,  
64, 72, 78, 79, 82, 91, 96, 97, 101n8,  
113–117, 120, 123, 124, 125, 128,  
129n6, 131, 133, 134, 135, 137, 140,  
141, 142, 147, 148, 150, 156–159,  
174; Argument from Intentionality  
128; Doubt Argument 114, 115;  
epistemological 133–136, 138,  
165; Extendedness and Divisibility  
Argument 115; modal 133–138, 165;  
semantic 133–137, 165  
Aristotle 13, 15, 19, 21, 22, 23, 27, 28, 33,  
43–47, 49, 51, 52, 53, 55, 56, 58, 60, 67,  
73–76, 95, 103, 109, 110, 111, 112, 113,  
131–139, 147, 165, 168, 173, 177  
Armstrong, D. M. 30n2, 31n11, 177  
article 29, 166; definite 11;  
indefinite 11, 69  
atheism 116  
Austin, J. L. 121, 122, 125, 128, 177  
axiom 68, 78, 79, 82; Axiom of  
Extensionality 35, 68, 69; Playfair's  
Axiom 80; Triangle Axiom 80  
Ayer, A. J. 21, 122, 150, 177  
  
'be' 19; of existence 19; of identity 19; of  
predication 19  
belief 69, 101, 103, 104, 106, 107, 108,  
111, 116, 118, 119, 120, 128, 131, 144,  
147–154, 166, 167n1, 168n12, 169n13,  
169n14, 178; basic (or foundational)  
149, 150; de dicto 112; de re 112;  
non-basic 149, 150; perceptual  
151  
Berkeley, G. 4, 69, 116, 177  
biconditional 6, 14, 21, 47, 64, 65,  
71, 95, 96  
Black, M. 32n16, 177  
Boghossian, P. A. 99, 177  
Bolyai, J. 80  
BonJour, L. 150, 169, 177  
brain in a vat 116, 117, 131, 147, 156, 157,  
158, 165  
British Empiricist 69  
Broadbent, A. 100n4, 177  
Burge, T. 140, 177  
  
Cantor, G. 34  
Carnap, R. 43, 63n10, 113, 177  
Carroll, L. 28, 177  
Casati, R., and Varzi, A. C. 61, 177  
causal 120, 136, 137, 140, 142, 156, 161,  
162; chain 136, 137; descriptivism 137,  
142, 178; efficacy 119, 120; theory 132,  
135, 136, 140, 156, 165, 166; *see also*   
 on

- causation 21, 85, 88, 100, 178, 180;  
     constant conjunction (or regularity)  
     theory of 88, 96; counterfactual analysis  
     of 88, 89
- Cavendish, H. 83
- Chalmers, D. 85, 129n3, 177
- Chisholm, R. 151, 177
- Churchland, P. M. 108, 177
- closeness 87, 96
- cluster concept theory 133, 136,  
     138, 168n5
- coarse-grained 47, 84
- cognitive process 153, 154; reliable 153,  
     169n13, 169n14; unreliable 153
- coherentism 129n1, 131, 147, 149, 150,  
     151, 152, 154, 165, 166
- complement 36, 38, 59
- Compositionality Principle 45, 47, 59, 60,  
     123; for Extension 45, 109, 111, 113;  
     for Intension 45
- concept 14, 18, 25, 32n16, 36, 38, 48, 49,  
     55, 58, 59, 64, 65, 68, 69, 70, 73, 75, 77,  
     78, 84, 85, 87, 89, 91, 93, 95, 96, 98,  
     101n8, 103, 121, 126, 130–133, 136,  
     138, 139, 142, 144, 145, 151, 152, 154,  
     165, 168; individual 43, 44, 46, 51, 54,  
     55, 59, 111, 132
- concrete object 2
- condition 20, 21, 53, 72, 87, 88, 91, 101,  
     102n12, 108, 148, 152, 164; necessary  
     12, 20, 21, 26, 27, 53, 148, 153, 173;  
     necessary and sufficient 21, 27, 173;  
     sufficient 12, 20, 21, 26, 27, 148, 173;  
     *see also* identity; truth
- conditional 6, 7, 20, 38, 47, 65, 66, 72,  
     85, 86, 91, 96, 99, 123, 180; indicative  
     7, 85, 86, 87, 96; material 7, 85, 86, 87,  
     96; subjunctive (or counterfactual) 7,  
     85–89, 92, 96
- conjunction 6, 30, 38, 65, 88, 96
- connective 6, 13, 66; logical 6, 65, 67
- connotation 42, 43, 46, 59
- consequent 6, 85, 155
- consistency 48, 91, 92
- consistent 42, 48, 63n10, 91, 113,  
     141, 146; formally 48; logically 18;  
     materially 48
- constitution 75; material 41, 100, 179;  
     view 75, 100, 146
- content 10, 69, 84, 103–107, 111, 113,  
     116, 119, 126, 127, 141, 142, 143, 148,  
     161, 162; de dicto mental 112, 126; de  
     re mental 112, 126; linguistic 103, 104,  
     105; mental 103–106, 108, 111, 113,  
     116, 118, 119, 126, 131, 140, 141, 142,  
     146, 166, 178; narrow 140, 142, 143,  
     144, 165; representational 104, 105,  
     118, 120–123, 125, 169n17; wide 140,  
     142, 143, 144, 165
- context 1, 6, 13, 22, 24, 35, 63n11, 110,  
     113, 136, 137; extensional 109, 110,  
     111, 115, 126, 128; intensional  
     109–114, 126, 128
- Continental Rationalist 69
- continuous 95, 102n15
- contradiction 67, 76, 92, 97, 157, 158
- cordate 34, 36, 38, 42, 43, 45, 51, 52, 54,  
     57, 61n1, 109, 111
- counterfactual 85–89, 92, 96, 178
- counterpart 49, 59
- Cresswell, M. J. 85, 178
- Davidson, D. 111, 144, 145, 146, 178
- Davies, M., and Humberstone, I. L.  
     168n6, 178
- de dicto 76, 97, 112, 113; *see also* belief;  
     desire; mental state; modality
- De Morgan's Law 7, 30n7; Generalized 9,  
     30n7, 71; Modal 71
- de re 76, 97, 112, 113; *see also* belief;  
     desire; mental state; modality
- demonstrative 10, 11, 32n18, 132
- denotation 21, 27, 31n10, 42, 43, 46, 47,  
     51, 52, 56, 59, 118, 122, 129n5, 133–  
     137, 168n5, 173
- dense 96, 102n15
- Descartes, R. 61n2, 69, 113, 114, 115, 116,  
     117, 129n3, 129n4, 149, 150, 159, 178
- description 55, 122, 125, 132–135, 137,  
     138, 139, 141, 142, 167n2, 168n5,  
     168n6, 178; definite 10, 11, 12, 19, 21,  
     26, 30n8, 44, 55, 56, 62n5, 62n8, 63n12,  
     63n13, 73, 132–138, 165, 167n2,  
     168n5, 168n6; indefinite 11, 19, 134;  
     parasitic 137, 138, 168n5; theory (or  
     descriptivism) 131–138, 165, 167n2
- descriptive fallacy 122
- desire 69, 103, 104, 106, 107, 116, 130; de  
     dicto 112; de re 112
- directedness 106
- Discourse on Method* 114, 178
- disjunction 6, 30, 38, 65, 66, 123, 133
- Disquotation Principle 157, 158, 165,  
     169n17; for Falsity 158

- domain 36, 38, 49, 50  
 Donnellan, K. 135, 178  
 dualism 113, 115, 126, 128, 129n3, 146;  
   property 114, 126, 145, 146; substance  
   113, 114, 126, 146  
 duplication problem 161, 162, 164, 165, 166
- Einstein, A. 81, 101n6, 134, 135,  
   165, 168n5  
*Elements* 79  
 emotivism 122, 126  
 endurantism 102  
 entity 2, 113, 115  
 epistemic 65, 77, 90, 177; certainty 72,  
   89; *see also* justification; logic; modality;  
   possible world  
 epistemology 48, 65, 91, 100, 117, 130,  
   131, 147, 156, 167, 180  
 equivalence 7, 14, 26, 101n10; class 14,  
   101n10; Equivalence Schema 157;  
   logical 14  
 equivalent 30, 38, 80, 157; logically  
   7, 14, 96  
 essence 4, 65, 72, 73, 74, 96, 98  
 Euclid 68, 78, 79  
 Evans, G. 99, 178  
 event 1, 88, 89, 131, 178  
 evil demon 114, 116, 117  
 exclusive 'or' 30n4  
 expression 6, 11, 21–25, 31n8, 43, 45,  
   46, 53, 54, 57, 58, 59, 62n3, 62n7, 76,  
   88, 102n16, 104, 109, 110, 111, 113,  
   114, 121, 122, 130, 132, 136, 154, 156;  
   linguistic 11, 12, 21, 42–45, 53, 60, 103,  
   104, 130  
 extension 6, 12, 34, 36, 38, 42–47, 51–60,  
   61n1, 62n7, 103, 106, 109, 110, 111,  
   113, 121; extensionless 110  
 extensional 42, 110, 111, 115, 126, 128;  
   co- 43, 45, 52, 57, 109–115; object 36,  
   38, 42, 46  
 externalism 91, 130, 131, 138, 140, 142,  
   147, 154, 155, 165, 166, 169n14, 178;  
   about epistemic justification 130, 131,  
   152, 165; about linguistic meaning  
   130; meaning 130; motivational 167n1;  
   reasons 130, 154; semantic 131, 136,  
   142, 147, 156, 157, 165
- fact 11, 13, 34, 39, 45, 56, 65, 67, 68,  
   69, 71, 72, 77, 78, 83, 86, 98, 102n16,  
   103, 104, 106, 109, 113, 115, 119, 120,  
   122, 130, 133, 134, 135, 152, 154, 161,  
   169n14, 175; moral 5, 118, 120, 121,  
   122, 123; natural 120; non-moral 120  
 falsity 12, 46, 47, 50, 63, 64, 87, 103, 106,  
   158; logical 67, 96, 174  
 Feynman, R. 134, 137  
 fictionalism 5; mathematical 5; modal 48  
 Fifth (or Parallel) Postulate 80  
 fine-grained 42, 57, 58  
 Fodor, J. 105, 178  
 form 11, 19, 79, 86, 88, 93, 103, 105, 106,  
   107, 114, 123, 126, 146, 147, 157  
 Form 4  
 formula 6, 103  
 foundationalism 131, 147, 149–152, 154,  
   165, 166  
 four-dimensionalism 16, 94  
 Frege, G. 6, 12, 13, 21, 43–46, 55, 56, 57,  
   61, 110, 111, 129n2, 129n6, 132, 133,  
   134, 178  
 Frege's Puzzle 44, 62n5, 83  
 Frege-Geach problem (or embedding  
   problem) 122, 123, 125, 126  
 function 49, 50, 51, 53, 54, 55, 58, 60,  
   63n11, 64, 85, 100, 107, 174; addition  
   49, 50; constant 50, 58, 72; cube 174;  
   division 50; doubling 49; mathematical  
   49, 59; multiplication 49, 50; partial  
   50, 55; square 49, 50; successor 49, 50;  
   total 50, 55
- Gauss, C. F. 81  
 Geach, P. 123, 178  
 General Relativity Theory 81, 101n6  
 generality problem 154  
 geometry 79, 81, 101n6; elliptic 80, 81,  
   82, 101n6; Euclidean 81, 82; hyperbolic  
   80, 81, 82; non-Euclidean 79, 97, 100  
 Gettier, E. 148, 178  
 Gettier problem 148, 165  
 Gödel, K. 134, 167n3  
 Goldman, A. 152, 178  
 Goliath 19, 74, 75, 98, 146, 147  
 grounding: problem of 75, 146, 147, 166
- haecceity 73, 96, 98, 138  
 Hare, R. M. 122, 178  
 Hawthorne, J. 29, 178  
 Hesperus 13, 15, 19, 44, 55, 56, 57, 70, 72,  
   82, 83, 84, 85, 92, 98, 105, 110, 138  
 human animal 19, 159, 164, 169, 179  
 Hume, D. 69, 88, 100, 116, 178

- idea 21, 30n1, 30n3, 30n8, 39, 43, 47, 48, 49, 68, 69, 74, 85, 92, 93, 95, 96, 100, 108, 116, 125, 139, 142, 145, 152, 161, 167
- identity 12–19, 29, 36, 38, 39, 43, 68, 75, 83, 110, 131, 159, 170n18; bodily 160; closest continuer theory of personal 170n19; condition 19, 47, 54, 133, 138; contingent 75, 96, 98; eternal 18; Identity of Indiscernibles 14, 32n16; necessary 75, 82, 96; numerical (or strict) 13–19, 26, 32n17, 75, 170; over time 17, 18, 95, 164; personal 18, 131, 158, 159–162, 166, 167, 169n18; predicate 12, 13, 19, 66; puzzle 32n17; psychological-continuity (or memory) theory of personal 159; relation 14; relative 18, 19, 26; statement 43, 44, 138; temporal (or diachronic) 17, 164; temporary 18, 75; theoretical 82, 83; token 15; transworld 49, 59; type 15
- immaterialism 116
- impossible world 92, 100, 102n12
- inclusive 'or' 30n4
- inconsistency 170n18
- inconsistent 48, 102n12, 169n18; formally 48; materially 48
- indexical 22, 48, 157
- Indiscernibility of Identicals 14, 114, 115, 128
- individual 2, 3, 4, 8, 11, 12, 15, 32, 35, 43, 44, 46, 49, 50, 51, 54, 55, 59, 62n7, 63n12, 73, 85, 95, 111, 121, 132, 138, 146, 151
- innate 84, 105
- intension 12, 34, 42–60, 62n7, 63n11, 65, 82–85, 103, 106, 109, 111, 121; primary 85; secondary 85
- intentional 34, 109–114, 126, 128; co- 43; object 34, 39, 42
- intentionality 115
- internalism 91, 130, 131, 138, 140, 142, 147, 154, 155, 165; about epistemic justification 130, 131, 152, 165; about linguistic meaning 130; about the meaning of natural kind terms 139; access 168n12; meaning 130; motivational 167n1; ontological 168n12; reasons 130, 154; semantic 131, 140, 165
- intersection 36, 38, 59
- introspection 150, 153, 154, 177
- Jackson, F. 85, 108, 178
- justification 99, 149, 150, 152, 154, 155, 169n14, 169n16; epistemic 129n1, 130, 131, 147, 149, 151, 152, 153, 161, 165
- Kant, I. 78, 79, 81, 82, 84, 97, 100, 100n5, 116, 178
- Kierkegaard, S. 5
- KK Principle 91, 96, 154, 155
- Kleene, S. C. 62n9, 178
- know-how 147, 168n10
- knowledge 65, 69, 84, 91, 99, 103, 108, 147–150, 153, 168n10, 177, 178, 179; a posteriori 69, 82, 83, 96; a priori 69, 77, 79, 84, 96; by acquaintance 168n10; JTB theory of 131, 147, 155, 165; mathematical 69, 84; propositional 147; theory of 65, 147
- know-that 147, 168n10
- Kripke, S. 13, 49, 55, 57, 61, 63n12, 73, 77, 78, 82, 83, 84, 93, 98, 99, 100, 101n8, 129n3, 132–140, 165, 167, 167n2, 178
- Kroon, F. 137, 178
- language 1, 5, 6, 8, 21, 24, 26, 27, 32, 46, 54, 56, 69, 83, 103, 104, 105, 107, 113, 117, 121, 126, 127, 130–133, 157, 167, 169n17, 177–180; meta 22, 24, 26, 157, 158; object 22, 24, 26, 157, 158; of thought 105, 126
- Language, Truth, and Logic* 21, 177
- LaPorte, J. 61, 178
- Lau, J., and Deutsch, M. 166, 178
- law 14, 114, 120; Law of Excluded Middle 66; Law of Non-Contradiction 67, 92; logical 9, 32n16; mathematical 48; *see also* De Morgan's Law; Leibniz's Law
- Lehrer, K. 150, 178
- Leibniz, G. W. 14, 25, 47, 69, 84
- Leibniz's Law 14, 32n14, 32n15, 114, 115, 128
- Lewis, D. 30n2, 48, 49, 87, 88, 92, 178
- Lobachevsky, N. 80, 81
- Locke, J. 18, 69, 101n7, 108, 159, 160, 161, 178
- Locke's theory: of perception 108, 116; of personal identity 131, 158–162, 164, 169n18

- logic 1, 5, 6, 7, 10, 12, 13, 20, 24, 29, 30n4, 34, 38, 47, 62, 65, 71, 85, 89, 92, 102n11, 124, 129, 177, 178, 179; classical 92; conditional 99; deontic 89, 96, 99; epistemic 89, 96, 99; first-order 6, 32n14; higher-order 32n14; many-valued 99; modal 49, 70, 71, 93, 96, 99, 168n6; non-classical 92, 99, 129; predicate 5, 6, 8, 11, 12, 13, 21, 27, 31n8, 32n14, 65; propositional (or sentential) 5, 8, 26, 30n7, 65, 86; quantificational 6, 9, 32n14; second-order 32n14; temporal (or tense) 95, 96, 97, 99
- Loux, M. J. 100, 178
- Lumpl 74, 75, 98, 146, 147
- Machery, E., Mallon, R., Nichols, S., and Stich, S. P. 167n4, 179
- Mackie, J. L. 118, 128, 179
- Marshall, D., and Weatherston, B. 166, 179
- Martinich, A. P., and Sosa, D. 167, 179
- material object 4, 16, 40, 72, 94
- mathematical object 5, 30n1, 54, 84, 138
- mathematics 13, 32, 34, 38, 41, 61, 70, 94, 100n1; applied 79, 82; pure 79, 82
- maximality 91, 102n12
- McLaughlin, B., and Bennett, K. 166, 179
- Meditation* 178; *First* 116, 129n4; *Second* 114, 129n4; *Sixth* 115
- Mellor, D. H., and Oliver, A. 29, 179
- memory 19, 69, 131, 150, 153, 154, 159, 161, 162, 164; apparent 161, 162, 163; genuine 160–163, 165; link 159, 160, 161; pseudo 160, 161, 162, 165
- mental state 69, 103, 104, 105, 107, 108, 122, 126, 131, 139, 140, 142, 143, 144, 150, 152, 153, 168n7, 168n12; de dicto 113; intentional 106; non-doxastic 150, 151, 152, 166; non-representational 108, 127; representational 106
- mentalese 105
- mereological 61n3, 62n3; atom 62n4; composition 41, 42; nihilism 62n4; part 40, 94; sum (or fusion) 39, 40, 41, 59, 73, 141; universalism 41, 62n4
- mereology 34, 38, 39, 41, 42, 60, 61, 61n3, 94
- metaphysics 4, 29, 33n19, 43, 45, 48, 55, 65, 76, 109, 111, 112, 130, 133, 178, 179, 180
- Mill, J. S. 63n12, 179
- Millian name 63n12
- mind 4, 48, 78, 88, 101n7, 103–108, 110, 113–117, 128, 130, 131, 132, 134–137, 139, 141, 147, 152, 154, 155, 159, 160, 162, 167n1, 169n17, 177, 179, 180; body problem 113, 126, 145
- modality 48, 70, 71, 72, 76, 77, 85, 89, 101n10, 113; de dicto 75, 76, 96, 112; de re 75, 76, 96; deontic 72, 85, 89, 90, 92, 96; epistemic 72, 85, 89, 90–93, 96, 113; metaphysical 72, 89, 91, 93, 96; temporal 85, 89, 94, 96
- mode of presentation 45
- moral 4, 5, 72, 118, 119, 120, 121, 122, 129n5, 144, 145, 146, 167, 178; discourse 121, 122, 123; error theory 118, 120, 126, 128; expressivism (or moral non-cognitivism) 121, 122, 123; judgment 116; non- 118, 120, 123, 125, 145; obligation 89, 93; permissibility 89, 93, 116, 118, 119, 120, 122, 129n5, 145; sentence 123, 124, 125; statement 122, 123, 125; theory 89; *see also* fact; property; realism
- morality 4, 116, 118, 120
- Musgrave, A. 100, 179
- Nagel, T. 108, 179
- Naming and Necessity* 55, 61, 99, 167, 178
- natural kind 139, 140, 165, 168n8; term 131, 139, 140, 141, 142
- necessary 12, 20, 21, 24, 26, 27, 48, 53, 64, 65, 68, 70–78, 82, 83, 84, 89, 91, 96, 98, 99, 100, 101n7, 101n8, 112, 137, 138, 148, 151, 153, 173; connection 88; existent (or being) 84; *see also* a posteriori; condition; identity; proposition; truth
- necessity 48, 55, 61, 64, 65, 68, 70, 71, 72, 78, 89, 90, 93, 99, 100n2, 167n3, 177, 178; epistemic 90
- negation 6, 38, 62n9, 63n9, 65, 66, 67, 102n12, 123, 129n5; external 63n9, 129n5; internal 63n9, 129n5
- Nelson, M. 128, 179
- Ney, A. 29, 179
- Nietzsche, F. 5
- nihilism 5, 62
- no competitor theory 162–165, 170n19
- Nolan, D. 100, 179
- Nolt, J., Rohatyn, D., and Varzi, A. 29, 179

- nominalism 4, 5, 26  
 Noumena 79, 116  
 noun 24, 98; abstract 31n12, 62n7;  
   common 3, 10, 11, 62n3, 98; phrase 11;  
   proper 11  
 Nozick, R. 170n19, 179  
 number 1, 4, 5, 10, 12, 13, 30n1, 34,  
   43, 50, 61, 84, 98, 102n15, 133, 136,  
   138, 139  
  
 Olson, E. T. 164, 179  
 omniscient 92  
 “On Denoting” 21, 179  
 ontology 4  
 operator 90; ‘actually’ 168n6; logical 13,  
   66; modal 92, 95, 112; necessity 90, 93;  
   possibility 93  
  
 part 3–6, 11, 16, 18, 19, 21, 31n9, 32n17,  
   35, 38–41, 45, 47, 59, 60, 61, 61n3, 78,  
   80, 81, 84, 86, 87, 88, 92, 94, 102n13,  
   109, 110, 111, 118, 121, 122, 124,  
   137, 150, 155, 162, 170n18, 177; body  
   127; grammatical 62n3; improper 40;  
   proper 40  
 particular 1–5, 8, 11, 12, 15, 26, 30n2,  
   30n3, 31n12, 32n14, 41, 95, 100n5;  
   abstract 30n1  
 perception 69, 108, 114, 116, 117, 120,  
   126, 150, 153, 166, 167n2, 168n5  
 perdurantism 102n13  
 performative utterance 120, 121, 122,  
   125, 126  
 Perry, J. 102n12, 167, 177, 179  
 persistence 17  
 person 4, 13, 18, 19, 51, 73, 122, 125, 134,  
   135, 159, 160, 162–165, 169n18  
 Phenomena 79  
 Phosphorus 13, 15, 19, 44, 55, 56, 57,  
   82–85, 92, 98, 105, 110, 138  
 physicalism (or materialism) 113, 126;  
   substance 146  
 Plantinga, A. 63n10, 168n6, 179  
 Plato 4, 13, 15, 22, 23, 26, 27, 28, 43, 44,  
   45, 51, 52, 55, 56, 62n5, 76, 103, 104,  
   105, 109, 111, 112, 132, 133, 134, 144,  
   148, 173  
 Plato’s Theory of Forms 4  
 platonism 4, 5, 12, 26  
 possibility 30n1, 48, 64, 65, 70, 71, 72,  
   89, 90, 93, 100n2, 107, 116, 125, 164,  
   167n1, 167n2, 169n16  
 possible world 25, 47–59, 63n10, 63n11,  
   64, 66, 70–76, 78, 82, 84, 87, 89, 90, 91,  
   92, 93, 95, 96, 98, 100, 101n7, 101n10,  
   102n12, 107, 113, 117, 120, 124, 134,  
   138, 168n6; conceptual 113; deontic 93;  
   epistemic 90, 91, 113; metaphysically  
   89, 101n10; possible worlds analysis  
   of counterfactuals 87; possible worlds  
   analysis of intension 49, 51, 53, 54,  
   57, 58, 60; possible worlds analysis of  
   modality 70; possible worlds semantics  
   34, 47, 49, 58, 82, 84, 85, 89, 96  
 predicate 5, 6, 8, 9, 10, 11, 12, 13, 21, 24,  
   27, 28, 29, 31n8, 31n9, 31n11, 31n12,  
   32n14, 34, 36, 38, 42, 43, 44, 46, 47,  
   49–54, 57, 58, 62n3, 62n7, 63n9, 65, 66,  
   85, 129n5, 157, 169n17, 170n18, 173;  
   0-place 8, 9, 10, 12, 26, 46; 1-place 8, 9,  
   10, 11, 12, 26, 42, 50, 60; 2-place 8, 9,  
   10, 12, 19, 26, 50; 3-place 8, 9, 11, 26,  
   50; adjectival 10, 26; adverbial 10, 19;  
   many-place 8, 12; nominal 10, 19, 26;  
   *n*-place 8, 9, 11; verbal 10, 26  
 prescriptivism 122, 126  
 Priest, G. 99, 102n14, 179  
 Principle of Unrestricted Mereological  
   Composition 41  
 product 36, 59  
 pronoun 22, 32  
 property 3, 8, 12, 14, 15, 26, 28, 29, 31n11,  
   32n14, 32n16, 42, 46, 47, 49, 51, 52, 54,  
   57, 73, 74, 75, 85, 111, 114, 115, 118,  
   119, 126, 129n5, 143–146, 169n17, 173,  
   179; abundance theory of properties  
   31n11; accidental 31n11, 73, 74, 96;  
   categorical 145, 146; dispositional  
   168n9; essential 31n11, 73, 74, 96, 98,  
   139; extrinsic 31n11, 131, 142, 143,  
   144, 165, 166, 179; intrinsic 31n11,  
   131, 142, 143, 144, 165, 166, 179;  
   macrophysical 145; mental 31n11, 114,  
   145, 146; microphysical 145; modal  
   75, 145, 146, 168n9; moral 5, 118–121,  
   129n5, 145, 146; natural 145, 146; non-  
   modal (or categorical) 145, 146, 147;  
   non-moral 118, 145; physical 31n11,  
   114, 145, 146; relational 143; set-  
   theoretic 35; sparse theory of properties  
   31n11, 32n16, 129n5  
 proposition 6, 8, 12, 46, 47, 48, 52, 53, 54,  
   58, 59, 63n10, 64, 66, 68, 69, 70, 72, 77,  
   78, 79, 82–85, 89, 91, 99, 100, 100n5,

- 101n7, 101n8, 102n12, 104, 106, 107, 109, 111, 113, 126, 129n2, 130, 138, 143, 147, 148, 150, 180; (actually) false 65, 70; (actually) true 64, 65, 69, 70, 78, 98; a posteriori 138; a priori 77, 134; analytic (or analytically true) 68, 77; contingent 70, 71; impossible 70; logically true 68; necessary (or necessarily true) 70, 72, 78, 83; structured 85; synthetically true 68
- propositional attitude 62, 105–110, 113, 114, 115, 126, 127, 128, 150, 179; report 107, 109, 110, 112, 113, 128, 129n6, 179
- Putnam, H. 117, 131, 139–142, 147, 156, 157, 158, 165, 168n7, 179
- Pythagoras's Theorem 79
- quality 61n1, 108; phenomenal 108, 126, 127; primary 61n1, 108; secondary 108
- quantifier 9, 13, 32n14, 66, 179; existential 9, 19, 30n6, 30n7, 65, 71; universal 9, 30n6, 30n7, 65, 71
- Quine, W.V. O. 34, 98, 102, 112, 113, 127, 128, 179
- quotation 12, 21–25, 76, 88, 113, 157; direct 22, 26, 107; indirect 22, 26, 107; mark 22–25, 27, 113
- quote 23, 24, 25, 27, 157; finger (or air) 24; scare 22, 24, 25, 26, 77
- range (or co-domain) 49, 50
- Ravenscroft, I. 128, 179
- Rea, M. 100, 179
- realism 4, 5, 26; about the external world 4; ersatz 63n10; mathematical 5; modal 48, 63n10; moral 4, 145, 177; naive (or direct) 108, 126
- reason 18, 25, 34, 59, 74, 77, 78, 98, 102n14, 111, 113, 115, 118, 119, 120, 125, 130, 144, 148, 153, 154, 163, 169, 179; external 130, 131; internal 130
- reasoning 49, 69; deductive 153; inductive 153, 154
- reductio ad absurdum 157
- reference 43–46, 59, 61, 111, 131, 132, 135–138, 140, 142, 156, 168n5, 178, 180; indirect 111; parasitic 137; –borrowing 137
- reflexivity 14
- reflexive 91, 93
- Reid, T. 108, 169n18, 179
- relation 3, 8, 12, 14, 19, 21, 26, 30n1, 35, 38, 64, 69, 70, 71, 74, 91, 96, 101n7, 103, 108, 113, 128, 132, 137, 144, 150, 167n1; 1-place 143; 2-place 3, 8, 14, 16, 74, 143; 3-place 3, 8, 74; accessibility 89, 90, 91, 93, 95, 101n10; causal 142, 156, 162; equivalence 14, 26, 101n10; external 74; identity 14; internal 74; *n*-place 143; part-whole (or part-of) 38, 39, 41; reference 137, 156; set-membership 35, 60
- reliabilism 131, 147, 149, 151–154, 165, 169n14
- renate 34, 36, 38, 42, 45, 51, 52, 54, 57, 61n1, 109, 111
- representation 104, 108
- representational 104–108, 118, 120–123, 125, 126, 169n17; non- 105, 107, 108, 123, 126, 127, 175; theory of perception 108, 116, 126; *see also* content; mental state
- Riemann, B. 80
- rigid designator 55, 56, 57, 59, 61, 73, 82, 98, 103, 134, 137, 138, 168, 178
- rigidify 138
- Russell, Bertrand 6, 12, 21, 30n8, 31n8, 62n8, 132, 133, 134, 150, 167n2, 179
- Russell, Bruce 99, 179
- Russell's theory of definite descriptions 30n8, 62n8
- Searle, J. 106, 133, 136, 137, 138, 141, 142, 165, 167, 168n5, 180
- Sellars, W. 151, 178
- semantic 65, 77, 103, 104, 111, 120, 121, 131, 133–137, 140, 142, 147, 156, 157, 165; innocence 111; *see also* argument; externalism; internalism
- semantics 34, 47, 48, 49, 58, 65, 82, 84, 85, 89, 96, 103, 121, 126, 156, 177; two-dimensional 85, 177; *see also* possible world
- sensation 150
- sense-data 150
- sentence 6, 7, 8, 10, 11, 12, 14, 23, 25, 26, 27, 32n14, 42, 43, 45, 46, 47, 52, 53, 54, 57, 58, 60, 62n3, 62n7, 62n8, 63n9, 63n10, 63n11, 64–69, 72, 73, 76, 83–86, 88, 91, 96–99, 102n16, 103, 104, 105, 109–113, 115, 121–127, 134, 138, 156,

- 157, 158; analytically true (or analytic) 68, 69, 70, 78, 84, 98, 134, 138; atomic 6, 8, 26, 47, 62n9, 67, 123; complex (or compound) 6, 7, 47, 62n9, 86, 123, 124; declarative 6, 8, 9, 10, 11, 12, 26, 42, 46, 121, 122, 126; exclamatory 10, 26; imperative 10, 26; interrogative 10, 26; logically true 66, 68, 69; necessarily true 58; synthetically true (or synthetic) 84, 98; true 46, 64, 98, 102n16, 109
- sententialism 113, 126, 127, 128
- set 5, 9, 30n1, 32n16, 34–43, 46, 49–52, 54, 59, 61, 63n10, 63n11, 71, 73, 84, 145, 151, 152, 174, 180; base (or subvenient) 145; empty (or null) 35, 67–70, 129; improper sub 37; membership 35, 36, 38, 39, 40, 41, 60; proper sub 37, 59; sub 37, 38, 40, 52, 59, 61; theory 34, 35, 38, 41, 60, 61; unit 35, 59
- Ship of Theseus 15, 16, 17, 19, 102, 162
- singular term 8–12, 30, 42–47, 51–55, 57, 60, 62n5, 62n7, 66, 85, 112, 115, 129n5, 132
- situation 1, 54, 102n12, 111, 140, 146, 148, 162, 166, 177
- skepticism 100, 116, 117, 118, 126, 131, 147, 156; methodological 129
- soul 160
- space 41, 78, 80, 81, 82
- spacetime 2–5, 16, 30n1, 30n2, 34, 35, 39, 40, 48, 61n2, 74, 78, 94, 101n6
- speech act theory 121, 126, 128
- Spinoza, B. 69
- Stalnaker, R. 63n10, 82, 83, 87, 100, 180
- state of affairs 63n10, 86, 104, 118, 119, 120, 126, 128, 161
- Steup, M. 167, 180
- Stevenson, C. L. 122, 180
- substance 73, 113, 114, 126, 139, 140, 141, 146, 159, 168n7
- substitutable *salva veritate* 109, 110, 114
- sum 36, 40, 41, 42, 59, 73, 78–81, 138, 141
- supervenience 131, 142, 144, 145, 146, 165, 166, 179; global 146; local (or individualistic) 146
- supervenient 145, 146; base 145; set 145
- Suppes, P. 61, 180
- symmetric 162
- symmetry 14, 15
- synonymous 19, 43, 46, 52, 53, 56, 58, 76, 103
- syntax (or grammar) 103, 121, 126
- tabula rasa 101
- tautology 65
- temporal 16, 17, 18, 39, 85, 89, 94–97, 99, 102n15, 164; order 21; part 94, 102n13, 170n18; slice 94; *see also* logic; modality; world
- that-clause 107, 109–112
- thought 46, 58, 62n6, 103–108, 126, 129n2, 156, 163, 178; experiment 139, 140, 142; *see also* language
- Three Dialogues between Hylas and Philonous* 116, 177
- three-dimensionalism 16, 102
- time 3, 13, 15, 16, 17, 18, 32n13, 32n17, 44, 47, 74, 75, 78, 84, 94–97, 102, 114, 123, 131, 136, 139, 143, 144, 145, 159, 160, 161, 164, 169, 170, 175
- token 1, 2, 111, 132, 141; event 1; expression 62n3; word 1, 15, 23, 26, 131
- transitive 10, 36, 41, 91
- transitivity 14, 18, 101n11, 102n11, 169n18, 170n18
- Trinity Thesis 77, 78, 84, 98, 99
- trope 30n1
- truth 12, 21, 44, 46, 47, 50, 53, 55, 58, 63n11, 64, 67, 68, 70, 71, 72, 77, 78, 79, 81, 82, 87, 93, 103, 106, 109, 123, 129n1, 153, 157, 169n17, 177, 179; actual 70, 71; analytic 64, 65, 67, 68, 72, 73, 77, 78, 82, 96, 97, 98, 100n1, 103; a posteriori 64, 65, 69, 70, 77, 82, 84, 96, 98, 134; a priori 64, 65, 69, 70, 77, 78, 81, 82, 96, 97, 98; -apt 122, 124, 125; coherence theory of 129n1; conceptual 67, 68, 72; condition 53; contingent 64, 65, 70, 77, 96, 98, 99; correspondence theory of 103, 169n17; deflationary (or disquotational) theory of 129n1, 169n17; functionality 47; logical 44, 64, 65, 66, 67, 68, 96, 97, 123, 174; mathematical 67, 68, 72, 78, 84, 100n1, 138, 151; necessary 48, 64, 65, 70, 72, 73, 77, 78, 83, 84, 96, 98, 99, 101n7, 138; non-logical 44, 64, 65, 67, 77, 96, 97; possible 96; predicate 157, 169n17; propositional 64; sentence 64; synthetic 64, 65, 67, 68, 77, 78, 96, 97, 98, 103;



- table 6, 7, 26, 66, 85; value 6, 7, 12, 23, 26, 27, 46, 47, 50, 53, 54, 58, 63n11, 64, 66, 71, 72, 100, 102n12, 107, 109, 112, 115, 122, 124, 174
- Twin Earth 139, 140, 141, 168n7
- type 1, 2, 10, 11, 15; event 1; expression 62n3; /token distinction 1; word 1, 23, 26
- union 36, 38, 59
- universal 1–5, 8, 9, 11, 12, 15, 26, 29, 30n1, 30n2, 30n3, 30n6, 30n7, 31n12, 46, 62n7, 65, 71, 79, 115, 177; immanent 30n1; transcendent 30n1
- use/mention distinction 22, 23, 26
- valid 5, 6, 96, 97, 115, 124, 174
- validity 5, 26, 32n16, 96, 97
- value 49, 50, 51, 60, 111, 144, 174; extrinsic 144, 165; intrinsic 5, 144, 165; *see also* truth
- van Fraassen, B. C. 62n9, 180
- verb 19, 22, 42, 43, 105, 106; auxiliary 72; intransitive 10; transitive 10
- Williams, D. C. 30, 180
- wishful thinking 153
- world 4, 10, 21, 22, 23, 25, 31n11, 46–58, 65–72, 75–81, 87, 89, 90, 91, 92, 93, 95, 98, 101n11, 102n12, 103–108, 112, 113, 114, 116, 120, 126, 138, 141, 145, 146, 147, 156, 169n17, 175, 178; actual 47–58, 63n10, 65, 66, 74, 75, 77, 78, 82, 84, 87–95, 98, 101n7, 107, 138, 146; conceptual 110, 111, 115; external 4, 116, 117; material 116, 128, 129n4; mental 146; morally ideal 92; physical 116, 145, 146; temporal 94, 95, 96, 102n15; worldbound 49; *see also* possible world; impossible world