The Philosophy Major's Introduction to Philosophy: Concepts and Distinctions

Book · August 2020

DOI: 10.4324/9781003039136

Ken Akiba
Virginia Commonwealth University

Chapter 1: Particulars and Universals; Logic and Language

Section 1.1: Tokens and Types; Particulars and Universals

To begin, let’s take a look at the whiteboard:

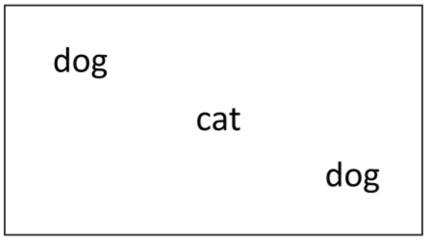


Figure 1.1: Words on the whiteboard

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. (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? How many animals are there?

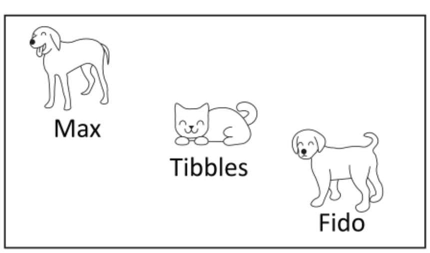


Figure 1.2: Animals in my backyard

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*.[1] 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.

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 contain 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*

[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 (1953, 1953a). 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.

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.[2]

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').[3] 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.

---

[2] Universals understood this way is 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.

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.

Section 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 19th 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.

Section 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*, ... (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)*,[4] *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.[5] 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*,

[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. *Not P* can be written as $\sim P$ or $-P$. *P and Q* can be written as $P \,\&\, Q$ or $P \cdot Q$. *If P then Q* and *P 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 only if Q* is equivalent to *Q if P*, which is equivalent to *if P then Q*.

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	**F**	T	T	T	F	T	**F**	F	T
T	F	**T**	T	F	F	F	T	**T**	T	F
F	T	**T**	F	F	T	T	F	**T**	F	T
F	F	**T**	F	F	F	T	F	**T**	T	F

P	Q	$\neg$	$(P$	$\vee$	$Q)$	$\neg$	P	$\wedge$	$\neg$	Q
T	T	**F**	T	T	T	F	T	**F**	F	T
T	F	**F**	T	T	F	F	T	**F**	T	F
F	T	**F**	F	T	T	T	F	**F**	F	T
F	F	**T**	F	F	F	T	F	**T**	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 $\neg(P \wedge Q)$ and $\neg P \vee \neg Q$ 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 tables show:

P	Q	P	$\rightarrow$	Q	$\neg$	P	$\vee$	Q
T	T	T	**T**	T	F	T	**T**	T
T	F	T	**F**	F	F	T	**F**	F

F	T	F	**T**	T	T	F	**T**	T
F	F	F	**T**	F	T	F	**T**	F

Section 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* (≥ 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.
- ¬*Tb*: Betty is not tall.
- *Lab*: Adam loves Betty.
- ¬*Lba*: Betty does not love Adam.
- *Lab* ∧ ¬*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*’).[6] However, we will not discuss them in details 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*).[7]

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? ....

---

[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 $Ta \wedge Tb \wedge Tc \wedge ...$ (Adam is tall and Betty is tall and Charlie is tall and ...), and $\exists xPx$ (something is a professor) means $Pa \vee Pb \vee Pc \vee ...$ (Adam is a professor or Betty is a professor or Charlie is a professor or ...). Because of that, a few textbooks use $\bigwedge x$ and $\bigvee x$ instead of $\forall x$ and $\exists x$ for the universal and the existential quantifier.

  - 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,[8] setting aside demonstratives. We will use the term '(linguistic) expression' as a more general term that includes singular terms, predicates, and sentences.

---

[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'

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, whomever 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)'.[9] 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[10] by '*a*' (Adam, Fido, etc.) instantiates the property denoted by '*Px*' (*being a professor*, *being a dog*, etc.).[11] The sentence is false if and only if 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 that in '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).

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[12]	stand for (represent, or express) ⇒	Particulars (individuals)
Predicates	⇒	Universals
– 1-place predicates	⇒	– Properties
– Many-place predicates	⇒	– Relations
(Declarative) sentences (= 0-place predicates)	⇒	– 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).

Section 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

[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.

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' means the same (identical) thing. But for a numerically identical thing, we tend to say 'one and the same[13] 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

[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 is 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'.)

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.[14]

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*.[15] 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.[16]

---

[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 I've 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.

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, water and H_2O, or the property of *being water* and that of *being dihydrogen monoxide*, are type identical; so everything that is water is dihydrogen monoxide and vice versa.

But, again, we do talk about identity that are 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

---

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.

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

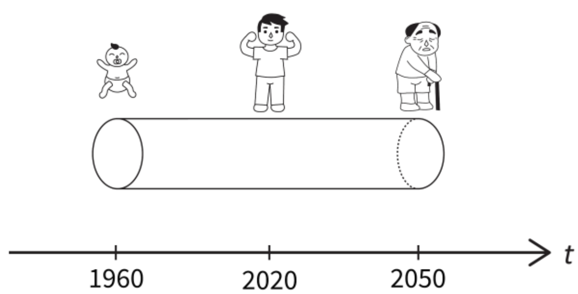


Figure 1.3: A four-dimensional object

Now, let's reconsider the Ship of Theseus puzzle on the basis of four-dimensionalism. Let's define the Original Ship of Theseus as the ship existing until time T1, when it loses its first plank, and the Repaired Ship as the ship beginning to exist at the harbor at T2 (>T1), 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 T2, when the last of the original planks are 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?

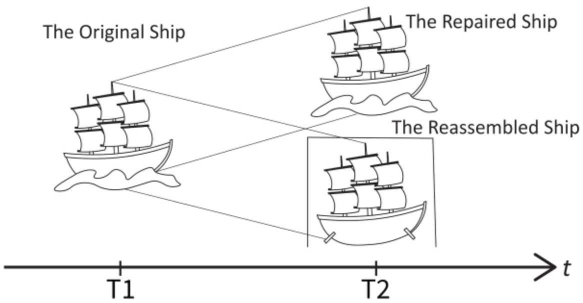


Figure 1.4: The Ship of Theseus

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 T1, the Repaired Ship and the Reassembled Ship do not exist until T2; 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 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.[17] 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*

[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.

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.

Section 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 precedes 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.

Section 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% 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 on 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:

Direct quotation	Indirect quotation
(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',[18] 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-lettered 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?)

[18] If these pronouns are used when you are directly pointing to objects, they are called 'demonstratives' (Section 1.4).

‘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, 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 below, which denotes the word token Adam produced:

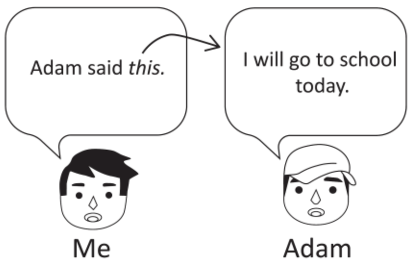


Figure 1.5: Adam said, ‘I will go to school today’

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-lettered 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 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.

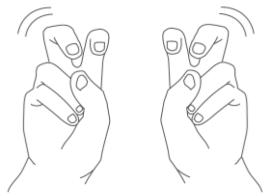


Figure 1.6: Finger quotes

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'.)[19] 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 is 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.

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'.

[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*').

– 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) $\mathrm{P} \leftrightarrow \neg\neg P$ (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 only has 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 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* (https://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.

Chapter 2: Extension and Intension

Section 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*,[1] 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.

Section 2.2: Set Theory

Set theory, first developed by the German mathematician Georg Cantor and formalized in the first half of the 20th 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:

[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.

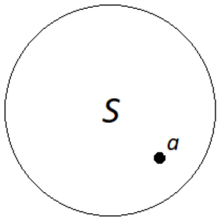


Figure 2.1: $a \in S$

Its denial is that *a* is not a member of *S* (or *a* is not in *S*), $a \notin S$:

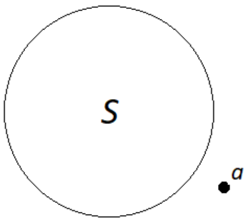


Figure 2.2: $a \notin S$

A set may be written like {*a*, *b*, *c*, …}; this means that *a*, *b*, *c*, … are (all) the members of the set. Also, a set may be written as, e.g., {*x*: *x* 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 $\varnothing$. 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 {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.

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* is a cordate} and the set of renates {*x*: *x* 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.[2]

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 $\overline{P}$), with respect to the domain *D* is *D* – *P*, where the

[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.

domain is the set of all things under consideration. So for any $x \in D$, $x \in -P$ iff $x \notin P$; or $-P = \{x \in D: x \notin P\}$.

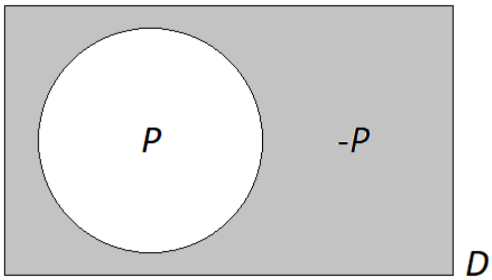


Figure 2.3: -*P*

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*. 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*.

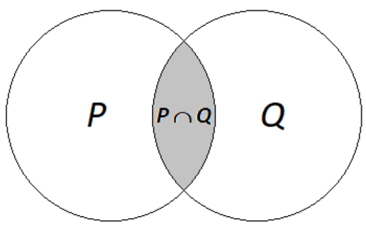


Figure 2.4: $P \cap Q$

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*. 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*.

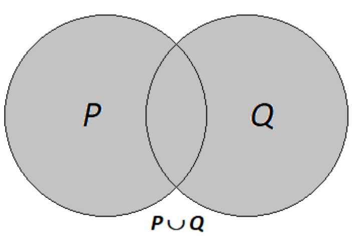


Figure 2.5: $P \cup Q$

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*. So, $P \subseteq Q$ iff for any *x*, if $x \in P$ then $x \in Q$.

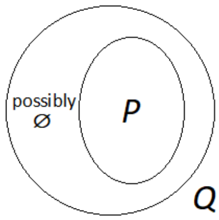


Figure 2.6: $P \subseteq Q$

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*. In other words, $P \subseteq Q$ iff $P \subset Q$ or $P = Q$.

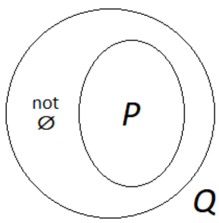


Figure 2.7: $P \subset Q$

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 are a proper subset of the set of Quakers.

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*, *-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*, $-P \cup Q$. 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:

- *-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 *-P*, $P \cap Q$, and $P \cup Q$ correspond to $\neg Px$, $Px \wedge Qx$, and $Px \vee Qx$, respectively. This is because *-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.

Section 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, all but 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[3] 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:

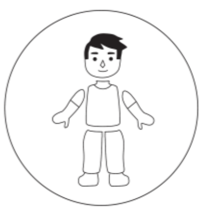


Figure 2.8: The set of my body parts

Since the identity of sets is determined solely by its members, this is the same set:

Figure 2.9: 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:

[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.

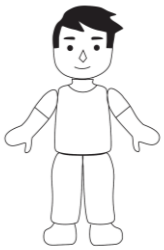


Figure 2.10: My body as a sum of my body parts

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 Frankensteinean 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.

Figure 2.11: A different sum of my body parts

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.

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 *b*. 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*.[4] 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.

Section 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 *in*volved in intension than the corresponding extension; so intensional objects, such as the properties *being a cordate* and *being a*

[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?

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 calls 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 are 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*.[5] 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.

[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.

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')[6] that the sentence expresses. From this line of reasoning, Frege concluded that the extension of a sentence is its truth value while the intension

[6] So Fregean thoughts exist even when nobody is holding those thoughts.

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[7]	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.,

[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.

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)[8] 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 its 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.[9] (In standard logic, however, we usually assume for the sake of simplicity that every sentence has a truth value.)

Section 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 ....

---

[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 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*.

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 are 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*.[10] 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.

[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.

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.

Section 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, 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.

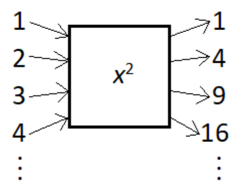


Figure 2.12: The square function

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 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.

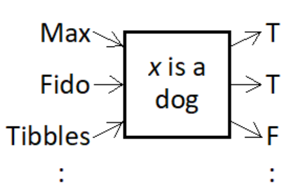


Figure 2:13: The '*x* is a dog' function

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 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.)

Section 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. 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.

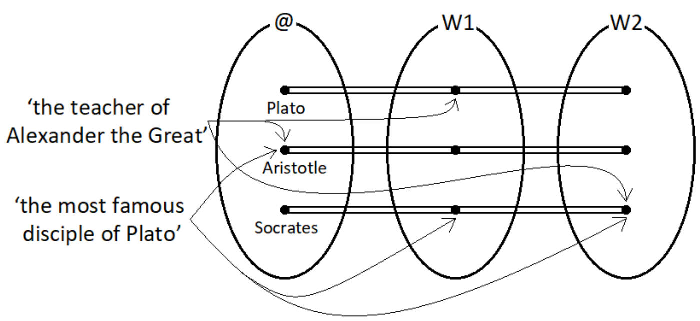


Figure 2.14: Possible denotations

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, 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 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.

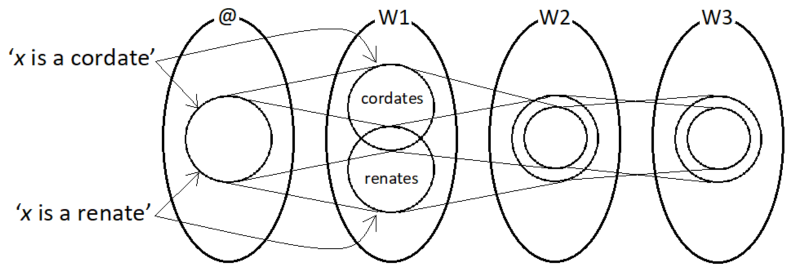


Figure 2.15: Possible extensions

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 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).

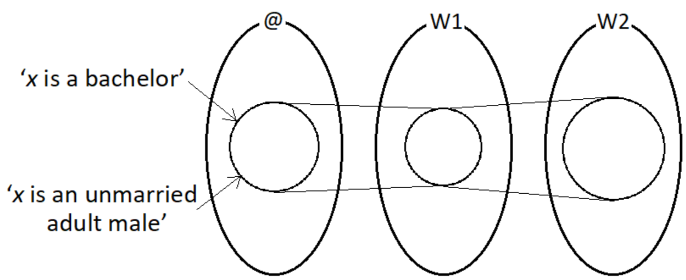


Figure 2.16: Synonymous predicates

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 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[11]

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 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.

[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.

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.

Section 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 long as Aristotle exists in those worlds. Similarly, 'Plato' rigidly denotes Plato and 'Socrates' rigidly denotes Socrates.[12]

[12] John Stuart Mill held a view similar to Kripke's:

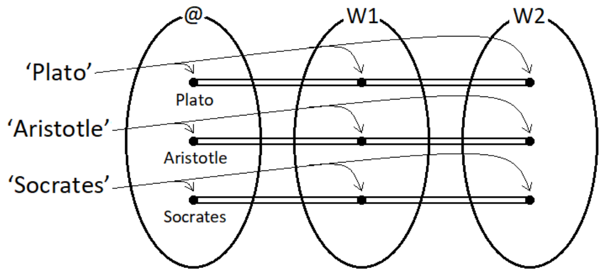


Figure 2.17: Rigid designators

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.[13]

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.

> 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.

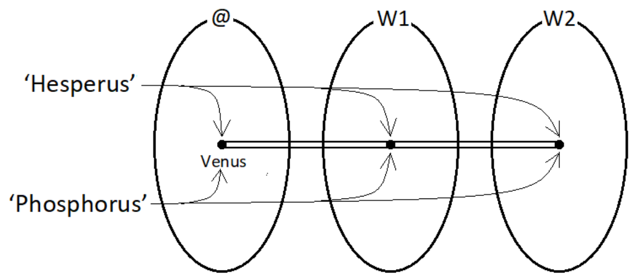


Figure 2.18: Rigid Designators co-extensional in the actual world

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

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

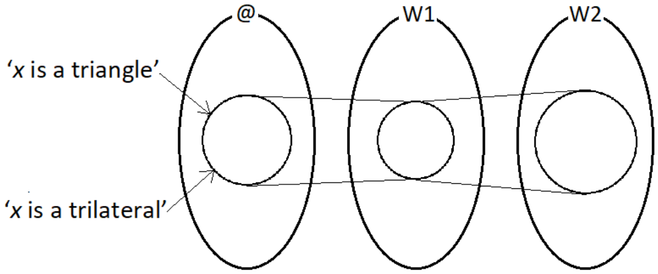


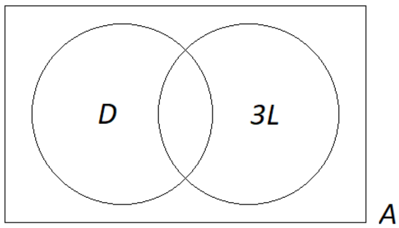
Figure 2.19: Synonymous predicates?

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 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

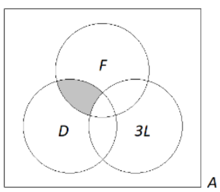
1. Explain the following concepts and distinctions.
   Sets/mereological sums (or fusions); extension (denotation, or reference)/intension (connotation, or sense); individual concepts; propositions; Compositionality Principle; possible worlds; transworld identity theory/counterpart theory; (mathematical) 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.



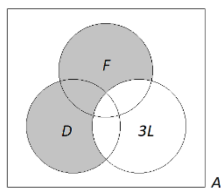
(a) The set of dogs.
(b) The set of animals except dogs.
(c) The set of three-legged dogs.
(d) The set of three-legged animals that are not dogs.
(e) The set of dogs and three-legged animals.
(f) The set of animals that are neither dogs nor three-legged.

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.)

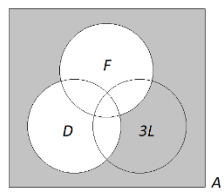
(g)



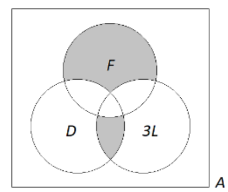
(h)



(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 → 3; 3 → 2.
   (b) 2 → 8; 4 → 64.
   (c) 9 → -18; -18 → 36.
   (d) Socrates → Plato; Plato → Aristotle.
   (e) Donald Trump → Barack Obama; Barack Obama → George W. Bush.
   (f) Aristotle → T; Donald Trump → 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 a fewer members than *S*. Discuss whether this claim is true or false. Hint: Compare the set *N* of natural numbers, 0, 1, 2, 3, ... with the set *E* of even natural numbers, 0, 2, 4, 6, ... *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.

Chapter 3: Analyticity, Apriority, and Necessity

Section 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 w* 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 tru*th 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.

Section 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	**T**	F	T
F	F	**T**	T	F

P	*P*	$\rightarrow$	*P*
T	T	**T**	T
F	F	**T**	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 $P \to P$ 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 '$a = a$', '$a = b \to b = a$', and '$(a = b \wedge b = c) \to a = c$' 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 $P \wedge \neg P$, $\neg(P \to P)$, 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.

Section 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.[1] 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

[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.

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.

Section 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 posterior’ 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.

Section 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.

<table>
<tr><td colspan="3">Possible</td><td rowspan="2">Impossible</td></tr>
<tr><td>Necessary</td><td colspan="2">Contingent</td></tr>
<tr><td colspan="2">Actually true</td><td colspan="2">Actually false</td></tr>
</table>

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 □*P* and ◊*P*, respectively, where □ and ◊ are called 'box' and 'diamond'.[2]

[2] Some textbooks use ◇ (tilted box) instead of ◊ for possibility. If you forget which symbol is for necessity and which is for possibility, just remember that the tilted box ◇ *can* fall down to the right or to the left (possibility), whereas the untilted box ☐ *must* stay that way (necessity).

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 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 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 is true in w) iff $\exists w\neg$(P 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.

Section 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 truth*s*'.

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[3] 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.

Section 3.7: Essence and Haecceity

[3] Following the convention in philosophy, throughout this book, when we talk about water, we will be including ice and steam.

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 hɛkˈsiːɪti, 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.

Section 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 Lumpl. The top half and the bottom half of Goliath are created separately and then put together to make Goliath; so Lumpl, being that specific single lump, was created at the same time as Goliath. Later, Goliath is crushed to smithereens; so Lumpl is destroyed at the same time as Goliath. Lumpl 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. Lumpl is indistinguishable from Goliath in its entire life. Is Lumpl 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 Lumpl is *being that single lump of clay*. We can think of a possible world in which the same Goliath exists but not Lumpl, for instance, Goliath loses one of its hands. We can also think of a possible world in which the same Lumpl exists but not Goliath, for instance, Lumpl is flattened. Lumpl 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 Lumpl 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 Lumpl 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 Lumpl may be contingently identical in the actual world but not numerically identical, according to this view.

Some philosophers disagree, however, and argue that Lumpl 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 Lumpl 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 Lumpl does not have; for it is equally puzzling that Goliath can have this property that Lumple does not have while having the same shape, size, temperature, etc., as Lumpl. 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.

Section 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.

Section 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:

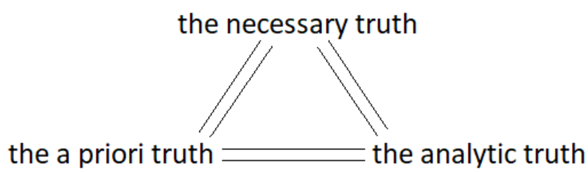


Figure 3.1: The Trinity Thesis

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.[4] 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.

[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.

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.

Section 3.11: 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°.[5] 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-19th century shortly after Kant's death, has revealed his mistakes. The story culminating the discovery is one of the most fascinating intellectual dramas in history, spanning over twenty centuries.

Euclid wrote thirteen Books of *Elements* around 300 BC. It was possibly the second most well-read book in western history only after the Bible. Until the 20th 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:

---

[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.

(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)

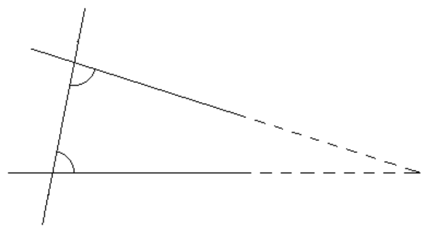


Figure 3.2: Parallel Postulate

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°. (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 in 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.

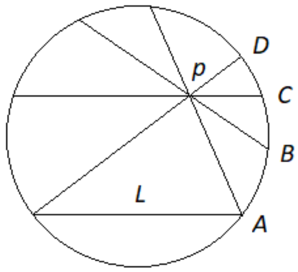


Figure 3.3: A model of two-dimensional hyperbolic geometry

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*, ..., 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.

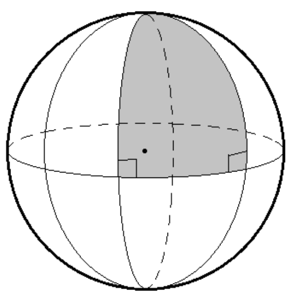


Figure 3.4: A model of two-dimensional elliptic geometry

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° (i.e., $\Delta > 180°$).

While it is true that the claim $\Delta = 180°$ is true in Euclidean geometry, a competing claim $\Delta < 180°$ is true in hyperbolic geometry, and another competing claim $\Delta > 180°$ is true in elliptic geometry. And which geometry captures the actual 'shape' of real, physical, space, and whether $\Delta = 180°$ or < 180° or > 180° 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, $\Delta > 180°$.[6] What Kant would have called an a priori truth, i.e., $\Delta = 180°$, has turned out to be not even true!

[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.

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.

Section 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 Hesperus =Phosphorus). Thus, that Hesperus = 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 Hesperus = Hesperus, it is necessary that Hesperus = Phosphorus. Therefore, that Hesperus = Phosphorus is a necessary but a posteriori truth.

Kripke used the same argument to show that many statements of theoretical identity, such as that water = H_2O and that heat = 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, Hesperus = Phosphorus is the same proposition as that if Hesperus exists, Hesperus = Hesperus, true in all possible worlds. Since we knew that if Hesperus exists, Hesperus = 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 give 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_2O, 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_2O', 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.[7]

---

[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

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.[8]

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

---

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 below.

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.

Section 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*;
- It it had been the case that *P*, then it would have been the case that *Q*.

We will symbolize these counterfactuals as '*P* > *Q*'.[9]

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 (1973; 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

[9] Other symbolizations are *P* □→ *Q* and *P* —﹥ *Q*.

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.

Section 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*. (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 (1973a) 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.

Section 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 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.[10]

[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.

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 @.

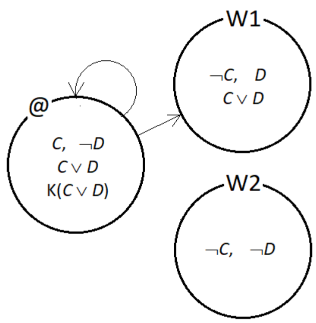


Figure 3.5: Epistemic possible worlds

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.)[11] 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)*.[12] 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

---

[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.

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 both *A* is 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 c*lassical 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., ¬*T* and *G*).

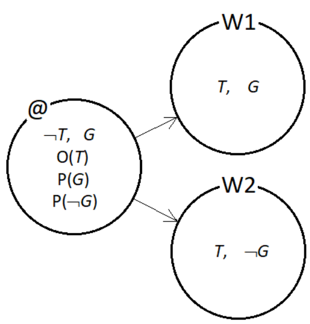


Figure 3.6: Deontic possible worlds

But the actual world has access to W1, in which *T* and *G*, as well as to W2, in which *T* and ¬*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:

- 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 world 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 ¬*G* is the case in W2, so both P(*G*) and P(¬*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 □*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?)

Section 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.[13] 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 exist in instantaneous slices of time. The whole picture looks like this:

[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* (& 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.

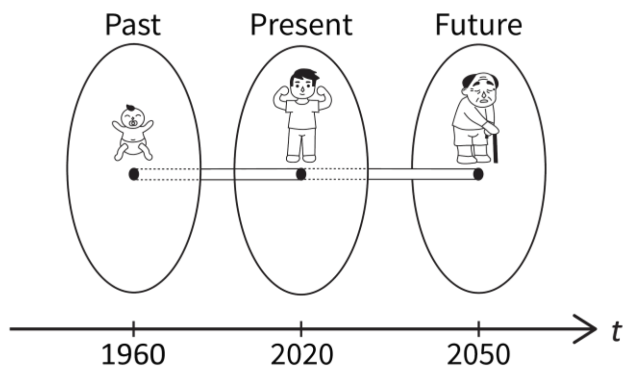


Figure 3.7: Temporal worlds

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 Now) 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 □ and ◊.[14] 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.

[14] For this reason, I love Priest's (2008) notation in which we write [P] and ⟨P⟩ instead of H and P, and [F] and ⟨F⟩ instead of G and F.

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*,[15] 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.

[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.

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.
   (a) The sky is blue or not blue.
   (b) All unmarried adult males are unmarried.
   (c) All bachelors are unmarried.
   (d) 21 is the legal minimum age for purchasing alcohol in the United States.
   (e) If 21 is the legal minimum age for purchasing alcohol in the United States, then 22 years olds can legally purchase alcohol in the United States.
   (f) 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.
   (g) If anyone at or over age 21 can legally purchase alcohol and you are 21 years-old, then you can legally purchase alcohol.
   (h) If I am happy only if I am rich, then if I am not rich, I am not happy.
   (i) All happy people are happy.
   (j) 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.
   (a) $\mathrm{H}(P \to Q) \to (\mathrm{H}(P) \to \mathrm{H}(Q))$
   (b) $\mathrm{G}(P \to Q) \to (\mathrm{F}(P) \to \mathrm{F}(Q))$
   (c) $\neg\mathrm{P}(P) \leftrightarrow \mathrm{H}(\neg P)$
   (d) $\neg\mathrm{G}(P) \leftrightarrow \mathrm{F}(\neg P)$
   (e) $(\mathrm{P}(P) \wedge \mathrm{P}(Q)) \to \mathrm{P}(P \wedge Q)$
   (f) $(\mathrm{F}(P) \vee \mathrm{F}(Q)) \to \mathrm{F}(P \vee Q)$
   (g) $P \to \mathrm{GP}(P)$
   (h) $\mathrm{HF}(P) \to (P \vee \mathrm{F}(P))$
   (i) $\mathrm{G}(P) \to \mathrm{F}(P)$

Discussion Questions

1. In his influential paper "Two Dogmas of Empiricism" (1951), W. V. 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.[16] 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 Lumpl 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, it 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 primateen' as 'a prime number larger than 12 but smaller than 20'. Then the sentence 'Every primateen 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:

[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 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.

(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 Scepticism*

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.