Chapter 2

What Is Philosophy?


“Two years!” said Dantès. “Do you think I could learn all this in two years?”
“In their application, no; but the principles, yes. Learning does not make one
learned: there are those who have knowledge and those who have understanding.
The first requires memory, the second philosophy.”
“But can’t one learn philosophy?”
“Philosophy cannot be taught. Philosophy is the union of all acquired knowl-
edge and the genius that applies it . . . ”
—Alexandre Dumas (1844, The Count of Monte Cristo, Ch. 17, pp. 168–169)

Philosophy is the microscope of thought.

Philosophy . . . works against confusion
http://www.publicphilosophy.org/resources.html#cleese

Consider majoring in philosophy. I did. . . . [I]t taught me how to break apart
arguments, how to ask the right questions
—NPR reporter Scott Simon, quoted in Keith 2014

To the person with the right turn of mind, . . . all thought becomes philosophy.
—Eric Schwitzgebel (2012).

Philosophy can be any damn thing you want!
—John Kearns (personal communication, 7 November 2013)
2.1 Readings

1. Very Strongly Recommended:

2. Strongly Recommended:
   - Plato, *The Apology* (various versions are online: search for “Plato Apology”) – Plato’s explanation of what Socrates thought that philosophy was all about; a good introduction to the skeptical, questioning nature of philosophy.

3. Recommended:
     (a) Ch. 3: “AI and the History of Philosophy” (pp. 19–40)
     (b) Ch. 4: “AI and the Rise of Contemporary Science and Philosophy” (pp. 41–50)
     – Some of the material may be online at the Google Books website for this book: http://tinyurl.com/Colburn00
2.2 Introduction

[W]e're all doing philosophy all the time. We can't escape the question of what matters and why: the way we're living is itself our implicit answer to that question. A large part of a philosophical training is to make those implicit answers explicit, and then to examine them rigorously. Philosophical reflection, once you get started in it, can seem endlessly demanding. But if we can't avoid living philosophically, it seems sensible to learn to do it well.

—David Egan (2019)

“What is philosophy?” is a question that is not a proper part of the philosophy of computer science. But, because many readers may not be familiar with philosophy, I want to begin our exploration with a brief introduction to how I think of philosophy, and how I would like non-philosophical readers who are primarily interested in computer science to think of it.

So, in this chapter, I will give you my definition of ‘philosophy’. We will also examine the principal methodology of philosophy: the evaluation of logical arguments (see §§2.6.1 and 2.10).

A Note on Quotation Marks:

Many philosophers have adopted a convention that single quotes are used to form the name of a word or expression. So, when I write this:

‘philosophy’

I am not talking about philosophy! Rather, I am talking about the 10-letter word spelled p-h-i-l-o-s-o-p-h-y. This use of single quotes enables us to distinguish between a thing that we are talking about and the name or description that we use to talk about the thing. This is the difference between a number (a thing that mathematicians talk about) and a numeral (a word or symbol that we use to talk about numbers). It is the difference between Paris (the capital of France) and ‘Paris’ (a 5-letter word). The technical term for this is the ‘use-mention distinction’ (http://en.wikipedia.org/wiki/Use-mention_distinction): We use ‘Paris’ to mention Paris. (For a real-life example, see §7.3.4.)

I will use double quotes when I am directly quoting someone. I will also sometimes use double quotes as “scare quotes”, to indicate that I am using an expression in a special or perhaps unusual way (as I just did). And I will use double quotes to indicate the meaning of a word or other expression.

2.3 A Definition of ‘Philosophy’

The word ‘philosophy’ has a few different meanings. When it is used informally, in everyday conversation, it can mean an “outlook”, as when someone asks you what your “philosophy of life” is. The word ‘philosophical’ can also mean something like “calm”, as when we say that someone takes bad news “very philosophically” (that is, very calmly).

But, in this chapter, I want to explicate the technical sense of modern, analytic, Western philosophy—a kind of philosophy that has been done since at least the time of
Socrates. ‘Modern philosophy’ is itself a technical term that usually refers to the kind of philosophy that has been done since René Descartes, who lived from 1596 to 1650, almost 400 years ago (Nagel, 2016). It is “analytic” in the sense that it is primarily concerned with the logical analysis of concepts (rather than literary, poetic, or speculative approaches). And it is “Western” in the sense that it has been done by philosophers working primarily in Europe (especially in Great Britain) and North America—though, of course, there are very many philosophers who do analytic philosophy in other areas of the world (and there are many other kinds of philosophy).

Further Reading:

On non-Western philosophy, consider this observation:

… there are good reasons to doubt that Greece, India, and China were the only societies that practiced philosophy, indeed to doubt that philosophy needed to be born or “invented” in the first place. Why not assume that philosophy is just a universal aspect of human culture? To explore this hypothesis, we need some idea of what it means for thoughts to be “philosophical.” This is a notoriously difficult question to answer, though most people probably feel that philosophy is like pornography: we know it when we see it. Provisionally, we might agree to apply the term to all abstract reflection on deep questions concerning ethics, knowledge, being, language, and so on. If that is what we are looking for, then perhaps we will find philosophy just about everywhere. (Adamson, 2019).

Western philosophy began in ancient Greece. Socrates (470–399 B.C.E.,1 that is, around 2500 years ago) was opposed to the Sophists, a group of teachers who can be caricatured as an ancient Greek version of “ambulance-chasing” lawyers, “purveyors of rhetorical tricks” (McGinn, 2012b). The Sophists were willing to teach anything (whether it was true or not) to anyone, or to argue anyone’s cause (whether their cause was just or not), for a fee.

Like the Sophists, Socrates also wanted to teach and argue, but only to seek wisdom: truth in any field. In fact, the word ‘philosophy’ comes from Greek roots meaning “love of [philo] wisdom [sophia]”. The reason that Socrates only sought wisdom rather than claiming that he had it (like the Sophists did) was that he believed that he didn’t have it: He claimed that he knew that he didn’t know anything (and that, therefore, he was actually wiser than those who claimed that they did know things but who really didn’t). As Victor Hugo put it, “the wise one knows that he is ignorant” (“Le savant sait qu’il ignore”; cited in O’Toole 2016), or, as the contemporary philosopher Kwame Anthony Appiah said, in reply to the question “How do you think Socrates would conduct himself at a panel discussion in Manhattan in 2019?”:

1 ‘B.C.E.’ is the abbreviation for ‘before the common era’; that is, B.C.E. years are the “negative” years before the year 1, which is known as the year 1 C.E. (for “common era”).
2.3. A DEFINITION OF ‘PHILOSOPHY’

You wouldn’t be able to get him to make an opening statement, because he would say, “I don’t know anything.” But as soon as anybody started saying anything, he’d be asking you to make your arguments clearer—he’d be challenging your assumptions. He’d want us to see that the standard stories we tell ourselves aren’t good enough. (Libbey and Appiah, 2019)

Socrates’s student Plato (430–347 B.C.E.), in his dialogue Apology, describes Socrates as playing the role of a “gadfly”, constantly questioning (and annoying!) people about the justifications for, and consistency among, their beliefs, in an effort to find out the truth for himself from those who considered themselves to be wise (but who really weren’t). (For a humorous take on this, see Figure 2.2.)

Plato defined ‘philosopher’ (and, by extension, ‘philosophy’) in Book V of his Republic (line 475c):

The one who feels no distaste in sampling every study, and who attacks the task of learning gladly and cannot get enough of it, we shall justly pronounce the lover of wisdom, the philosopher. (Plato, 1961b, p. 714, my emphasis).

Adapting this, I define ‘philosophy’ as:

the personal search for truth, in any field, by rational means.

This raises several questions:

1. Why only “personal”? (Why not “universal”?)
2. Why is philosophy only the search for truth? (Can’t we succeed in our search?)
3. What is “truth”?  
4. What does ‘any field’ mean?  
   (Is philosophy really the study of anything and everything?)
5. What counts as being “rational”?  

Let’s look at each of these, beginning with the second.
2.4 What Is Truth?

The study of the nature of truth is one of the “Big Questions” of philosophy, along with things like: What is the meaning of life? What is good? What is beauty? and so on.

I cannot hope to do justice to it here, but there are two theories of truth that will prove useful to keep in mind on our journey through the philosophy of computer science: the correspondence theory of truth and the coherence theory of truth.

Further Reading:
On “the Big Questions”, see §2.8, below, and Gabriel Segal’s response to the question “What is it that is unique to philosophy that distinguishes it from other disciplines?”, http://www.askphilosophers.org/question/5017.

2.4.1 The Correspondence Theory of Truth

The correspondence theory states that a belief is true if and only if that belief corresponds to the facts. . . . It captures the idea that truth depends on objective reality—not on us. The problem the correspondence theory has concerns more technical issues such as what a fact is and what the correspondence relation amounts to.
—Colin McGinn (2015a, pp. 148–149)

The word ‘true’ originally meant “faithful”. Such faithfulness requires two things A and B such that A is faithful to B. According to the correspondence theory (see David 2009), truth is faithfulness of (A) a description of some part of reality to (B) the reality that it is a description of. On the one hand, there are beliefs (or propositions, or sentences); on the other hand, there is “reality”: A belief (or a proposition, or a sentence) is true if and only if (“iff”) it corresponds to reality, that is, iff it is faithful to, or “matches”, or accurately characterizes or describes reality.

Terminological Digression and Further Reading:
A “belief”, as I am using that term here, is a mental entity, “implemented” (in humans) by certain neuron firings. A “sentence” is a grammatical string of words in some language. And a “proposition” is the meaning of a sentence. These are all rough-and-ready characterizations; each of these terms has been the subject of much philosophical analysis. For further discussion, see Schwitzgebel 2015 on belief, https://en.wikipedia.org/wiki/Sentence_(linguistics) on sentences, and King 2016 on propositions.

To take a classic example, the three-word English sentence ‘Snow is white.’ is true iff the stuff in the real world that precipitates in certain winter weather (that is, snow) has the same color as milk (that is, iff it is white). Put somewhat paradoxically (but correctly—recall the use-mention distinction!), ‘Snow is white.’ is true iff snow is white.
2.4. WHAT IS TRUTH?

Further Reading:
The standard logical presentation of a correspondence theory of truth is due to Alfred Tarski. See Hodges 2018 for an overview and further references, and Tarski 1969 for a version aimed at a general audience.

How do we determine whether a sentence (or a belief, or a proposition) is true? On the correspondence theory, in principle, we would have to compare the parts of the sentence (its words plus its grammatical structure, and maybe even the context in which it is thought, uttered, or written) with parts of reality, to see if they correspond. But how do we access “reality”? How can we do the “pattern matching” between our beliefs and reality?

One answer is by sense perception (perhaps together with our beliefs about what we perceive). But sense perception is notoriously unreliable (think about optical illusions, for instance). And one of the issues in deciding whether our beliefs are true is deciding whether our perceptions are accurate (that is, whether they match reality).

So we seem to be back to square one, which gives rise to the coherence theory.

2.4.2 The Coherence Theory of Truth

The coherence theory states that a proposition is true if and only if that proposition coheres with the other propositions that one believes. . . . The problem with the coherence theory is that a belief could be consistent with my other beliefs and yet the whole lot could be false.

—Colin McGinn (2015a, p. 148)

According to the coherence theory of truth (see Young 2018), a set of propositions (or beliefs, or sentences) is true iff:

1. they are mutually consistent, and

2. they are supported by, or consistent with, all available evidence;

that is, they “cohere” with each other and with all evidence.

Note that observation statements (that is, descriptions of what we observe in the world around us) are among the claims that must be mutually consistent, so this is not (necessarily) a “pie-in-the-sky” theory that doesn’t have to relate to the way things really are. It just says that we don’t have to have independent access to “reality” in order to determine truth.

2.4.3 Correspondence vs. Coherence

Which theory is correct? Well, for one thing, there are more than two theories: There are several versions of each kind of theory, and there are other theories of truth that don’t fall under either category. The most important of the other theories is the “pragmatic” theory of truth (see Glanzberg 2016, §3; Misak and Talisse 2019). Here is one version:
[The] pragmatic theory of truth . . . is that a proposition is true if and only if it is useful [that is, “pragmatic”, or practical] to believe that proposition. (McGinn, 2015a, p. 148)

Another version states that a belief, proposition, or sentence is true iff it continues to be accepted at the limit of inquiry:

Truth is that to which a belief would tend were it to tend indefinitely to a fixed belief. (Edwin Martin, Jr., paraphrasing C.S. Peirce; lectures on the theory of knowledge, Indiana University, Spring 1973; for more on Peirce, see §2.6.1.3, below.)

However, “I could have a belief about something that is useful to me but that belief is false” (McGinn, 2015a, p. 149). Similarly, a “fixed” belief that remains “at the limit of inquiry” might still be false.

Fortunately, the answer to which kind of theory is correct (that is, which kind of theory is, if you will excuse the expression, true) is beyond our present scope! But note that the propositions that a correspondence theory says are true must be mutually consistent (if “reality” is consistent!), and they must be supported by all available evidence; that is, a correspondence theory must “cohere”. Moreover, if you include both propositions and “reality” in one large, highly interconnected network, that network must also “cohere”, so the propositions that are true according to a coherence theory of truth should “correspond to” (that is, cohere with) reality.

Let’s return to the question raised in §2.4.1, above: How can we decide whether a statement is true? One way that we can determine its truth is syntactically (that is, in terms of its grammatical structure only, not in terms of what it means), by trying to prove it from axioms via rules of inference. It is important to keep in mind that, when you prove a statement this way, you are not proving that it is true! You are simply proving that it follows logically from certain other statements, that is, that it “coheres” in a certain way with those statements. But, if the starting statements—the axioms—are true (note that I said “if they are true”; I haven’t told you how to determine their truth value yet), and if the rules of inference “preserve truth”, then the statement that you prove by means of them—the “theorem”—will also be true. (Briefly, rules of inference—which tell you how to infer a statement from other statements—are truth-preserving if the inferred statement cannot be false as long as the statements from which it is inferred are true.)

**Further Reading:**

I’ll say more about what axioms and rules of inference are in §§6.6, 7.6.5, 14.3.2.1, and 16.2. For now, just think of proving theorems in geometry or logic.

Another way we can determine whether a statement is true is semantically (that is, in terms of what it means). This, by the way, is the only way to determine whether an axiom is true, since, by definition, an axiom cannot be inferred from any other statements. (If it could be so inferred, then it would be those other statements that would be the real axioms.)
But to determine the truth of a statement semantically is also to use syntax: We semantically determine the truth value of a complex proposition by syntactic manipulation (truth tables) of its atomic constituents. (We can use truth tables to determine that axioms are true.) (For more on the nature of, and relation between, syntax and semantics, see §19.6.3.3.) How do we determine the truth value of an atomic proposition? By seeing if it corresponds to reality. But how do we do that? By comparing the proposition with reality, that is, by seeing if the proposition coheres with reality.

2.5 On Searching for the Truth vs. Finding It

Thinking is, or ought to be, a coolness and a calmness . . . .
—Herman Melville (1851, Moby-Dick, Ch. 135, p. 419)

Thinking is the hardest work there is, which is the probable reason why so few engage in it.
—Henry (Ford, 1928, p. 481)

Thinking does not guarantee that you will not make mistakes.
But not thinking guarantees that you will.
—Leslie Lamport (2015, p. 41)

How does one go about searching for the truth, for answering questions? As we’ll see below, there are basically two complementary methods: (1) thinking hard and (2) empirical investigation. We’ll look at the second of these in §2.6. In the present section, we’ll focus on thinking hard.

Some people have claimed that philosophy is just thinking really hard about things (see some of the quotes in Popova 2012). Such hard thinking requires “rethinking explicitly what we already believe implicitly” (Baars, 1997, p. 187). In other words, it’s more than just expressing one’s opinion unthinkingly. It’s also different from empirical investigation:

Philosophy is thinking hard about the most difficult problems that there are. And you might think scientists do that too, but there’s a certain kind of question whose difficulty can’t be resolved by getting more empirical evidence. It requires an untangling of presuppositions: figuring out that our thinking is being driven by ideas we didn’t even realize that we had. And that’s what philosophy is. (David Papineau, quoted in Edmonds and Warburton 2010, p. xx)

Can we find the truth? Not necessarily.

For one thing, we may not be able to find it. The philosopher Colin McGinn (1989, 1993) discusses the possibility that limitations of our (present) cognitive abilities may make it as impossible for us to understand the truth about certain things (such as the mind-body problem or the nature of consciousness) in the same way that, say, an ant’s cognitive limitations make it impossible for it to understand calculus.

But I also believe that finding it is not necessary; that is, we may not have to find it. Philosophy is the search for truth. Albert Einstein said that “the search for truth is more precious than its possession” (Einstein, 1940, p. 492, quoting G.E. Lessing). In a similar vein, the mathematician Carl Friedrich Gauss said, “It is not knowledge, but
the act of learning, not possession but the act of getting there, which grants the greatest enjoyment.”

Further Reading:
Here is Lessing’s (1778) original version of the Einstein quote:

The true value of a man [sic] is not determined by his possession, supposed or real, of Truth, but rather by his sincere exertion to get to the Truth. It is not possession of the Truth, but rather the pursuit of Truth by which he extends his powers . . . .

The Gauss quote is from his “Letter to Bolyai”, 1808, http://blog.gaiam.com/quotes/authors/karl-friedrich-gauss/21863

For more on the importance of search over success, see my website on William Perry’s theory of intellectual development, http://www.cse.buffalo.edu/~rapaport/perry-positions.html and Rapaport 1982. Perry’s theory is also discussed briefly in §2.7, below, and at more length in §C.

Digression:
The annotation ‘[sic]’ (which is Latin for “thus” or “so”) is used when an apparent error or odd usage of a word or phrase is to be blamed on the original author and not on the person (in this case, me!) who is quoting the author. For example, here I want to indicate that it is Lessing who said “the true value of a man”, where I would have said “the true value of a person”.

2.5.1 Asking “Why?”

Questions, questions. That’s the trouble with philosophy: you try and fix a problem to make your theory work, and a whole host of others then come along that you have to fix as well. —Helen Beebee (2017)

One reason that this search will never end (which is different from saying that it will not succeed) is that you can always ask “Why?”; that is, you can always continue inquiring. This is the way philosophy—and philosophers—are[:] Questions beget questions, and those questions beget another whole generation of questions. It’s questions all the way down. (Cathcart and Klein, 2007, p. 4)

You can even ask why “Why?” is the most important question (Everett, 2012, p. 38)! “The main concern of philosophy is to question and understand very common ideas that all of us use every day without thinking about them” (Nagel, 1987, p. 5). This is why, perhaps, the questions that children often ask (especially, “Why?”) are often deeply philosophical questions.

In fact, as the physicist John Wheeler has pointed out, the more questions you answer, the more questions you can ask: “We live on an island surrounded by a sea of ignorance. As our island of knowledge grows, so does the shore of our ignorance” (https://en.wikiquote.org/wiki/John_Archibald_Wheeler). And “Philosophy patrols the . . . [shore], trying to understand how we got there and to conceptualize our next move” (Soames, 2016). The US economist and social philosopher Thorstein Veblen said, “The
outcome of any serious research can only be to make two questions grow where only one grew before” (Veblen, 1908, p. 396).

Asking “Why?” is part—perhaps the principal part—of philosophy’s “general role of critically evaluating beliefs” (Colburn, 2000, p. 6) and “refusing to accept any platitudes or accepted wisdom without examining it” (Donna Dickenson, in Popova 2012).

Critical thinking in general, and philosophy in particular, “look . . . for crack[s] in the wall of doctrinaire [beliefs]—some area of surprise, uncertainty, that might then lead to thought” (Acocella, 2009, p. 71). Or, as the humorist George Carlin put it:

> [It’s] not important to get children to read. Children who wanna read are gonna read. Kids who want to learn to read [are] going to learn to read. [It's] much more important to teach children to QUESTION what they read. Children should be taught to question everything. ([http://www.georgecarlin.net/boguslist.html#question](http://www.georgecarlin.net/boguslist.html#question))

Whenever you have a question, either because you do not understand something or because you are surprised by it or unsure of it, you should begin to think carefully about it. And one of the best ways to do this is to ask “Why?”: Why did the author say that? Why does the author believe it? Why should I believe it? (We can call this “looking backward” towards reasons.) And a related set of questions are these: What are its implications? What else must be true if that were true? And should I believe those implications? (Call this “looking forward” to consequences.) Because we can always ask these backward- and forward-looking questions, we can understand why...

... Plato is the philosopher who teaches us that we should never rest assured that our view, no matter how well argued and reasoned, amounts to the final word on any matter. (Goldstein, 2014, p. 396)

This is why philosophy must be argumentative. It proceeds by way of arguments, and the arguments are argued over. Everything is aired in the bracing dialectic wind stirred by many clashing viewpoints. Only in this way can intuitions that have their source in societal or personal idiosyncrasies be exposed and questioned. (Goldstein, 2014, p. 39)

The arguments are argued over, typically, by challenging their assumptions. It is rare that a philosophical argument will be found to be invalid. The most interesting arguments are valid ones, so that the only concern is over the truth of its premises. An argument that is found to be invalid is usually a source of disappointment—unless the invalidity points to a missing premise or reveals a flaw in the very nature of logic itself (an even rarer, but not unknown, occurrence).

### 2.5.2 Can There Be Progress in Philosophy?

If the philosophical search for truth is a never-ending process, can we ever make any progress in philosophy? Mathematics and science, for example, are disciplines that not only search for the truth, but seem to find it; they seem to make progress in the sense that we know more mathematics and more science now than we did in the past. We have well-confirmed scientific theories, and we have well-established mathematical proofs of theorems. (The extent to which this may or may not be exactly the right way to look at things will be considered in Chapter 4.) But philosophy doesn’t seem to
be able to empirically confirm its theories or prove any theorems. So, is there any sense of “progress” in philosophy? Or are the problems that philosophers investigate unsolvable?

I think there can be, and is, progress in philosophy. Solutions to problems are never as neat as they seem to be in mathematics. In fact, they’re not even that neat in mathematics! This is because solutions to problems are always conditional; they are based on certain assumptions. Most mathematical theorems are expressed as conditional statements: If certain assumptions are made, or if certain conditions are satisfied, then such-and-such will be the case. In mathematics, those assumptions include axioms, but axioms can be challenged and modified: Consider the history of non-Euclidean geometry, which began by challenging and modifying the Euclidean axiom known as the Parallel Postulate.

Further Reading:
One version of the Parallel Postulate is this: For any line \( L \), and for any point \( P \) not on \( L \), there is only one line \( L' \) such that (1) \( P \) is on \( L' \), and (2) \( L' \) is parallel to \( L \). For some of the history of non-Euclidean geometries, see http://mathworld.wolfram.com/ParallelPostulate.html and http://en.wikipedia.org/wiki/Parallel_postulate

So, solutions are really parts of larger theories, which include the assumptions that the solution depends on, as well as other principles that follow from the solution. Progress can be made in philosophy (as in other disciplines), not only by following out the implications of your beliefs (“forward-looking” progress), but also by becoming aware of the assumptions that underlie your beliefs (“backward-looking” progress) (Rapaport, 1982):

Progress in philosophy consists, at least in part, in constantly bringing to light the covert presumptions that burrow their way deep down into our thinking, too deep down for us to even be aware of them. … But whatever the source of these presumptions of which we are oblivious, they must be brought to light and subjected to questioning. Such bringing to light is what philosophical progress often consists of. … (Goldstein, 2014, p. 38)

Philosophy is a “watchdog” (Colburn, 2000, p. 6). This zoological metaphor is related to Socrates’s view of the philosopher as “gadfly”, investigating the foundations of, or reasons for, beliefs and for the way things are, always asking “What is \( X \)?” and “Why?”. Of course, this got him in trouble: His claims to be ignorant were thought (probably correctly) to be somewhat disingenuous. As a result, he was tried, condemned to death, and executed. (For the details, read Plato’s Apology.)

One moral is that philosophy can be dangerous:

Thinking about the Big Questions is serious, difficult business. I tell my philosophy students: “If you like sweets and easy living and fun times and happiness, drop this course now. Philosophers are the hazmat handlers of the intellectual world. It is we who stare into the abyss, frequently going down into it to great depths. This isn’t a job for people who scare easily or even have a tendency to get nervous.” (Eric Dietrich, personal communication, 5 October 2006.)
And what is it, according to Plato, that philosophy is supposed to do? Nothing less than to render violence to our sense of ourselves and our world, our sense of ourselves in the world. (Goldstein, 2014, p. 40)

It is violent to have one’s assumptions challenged:

Philosophy is difficult because the questions are hard, and the answers are not obvious. We can only arrive at satisfactory answers by thinking as rigorously as we can with the strongest logical and analytical tools at our disposal.

… I want … [my students] to care more about things like truth, clear and rigorous thinking, and distinguishing the truly valuable from the specious.

The way to accomplish these goals is not by indoctrination. Indoctrination teaches you what to think; education teaches you how to think. Further, the only way to teach people how to think is to challenge them with new and often unsettling ideas and arguments.

… Some people fear that raising such questions and prompting students to think about them is a dangerous thing. They are right. As Socrates noted, once you start asking questions and arguing out the answers, you must follow the argument wherever it leads, and it might lead to answers that disturb people or contradict their ideology. (K.M. Parsons 2015)

So, the whole point of Western philosophy since Socrates has been to get people to think about their beliefs, to question and challenge them. It is not (necessarily) to come up with answers to difficult questions.

Further Reading:

Very similar comments have been made about science: “The best science often depends on asking the most basic questions, which are often the hardest to ask because they risk exposing fundamental limitations in our knowledge” (Mithen, 2016, p. 42).

For more on whether there can be progress in philosophy, see Rapaport 1982, 1984a; Rescher 1985; Moody 1986; Chalmers 2015; Frances 2017; as well as the answers to “Have philosophers ever produced anything in the way that scientists have?” and “How is ‘philosophical progress’ made, assuming it is made at all?”, at http://www.askphilosophers.org/question/2249 and http://www.askphilosophers.org/question/4523, respectively.

2.5.3 Skepticism

Sceptics do not always really intend to prove to us that we cannot know any of the things we naively think we know; sometimes they merely wish to demonstrate to us that we are too naïve about how we know them. … [S]ceptics have an uncanny eye for fundamental principles …


If you can always ask “Why?”—if you can challenge any claims—then you can be skeptical about everything. Does philosophy lead to skepticism?[^1]

[^1]: That’s the British spelling.
[^2]: See http://www.askphilosophers.org/questions/5572
Skepticism is often denigrated as being irrational. But there are advantages to always asking questions and being skeptical: “A skeptical approach to life leads to advances in all areas of the human condition; while a willingness to accept that which does not fit into the laws of our world represents a departure from the search for knowledge” (Dunning, 2007). Being skeptical doesn’t necessarily mean refraining from having any opinions or beliefs. But it does mean being willing to question anything and everything that you read or hear (or think!). Here is another way of putting this: In philosophy, the jury is always out!—see Polger 2011, p. 21. But, as we saw above, this does not mean that there can be no progress in philosophy.

Why would you want to question anything and everything? (See Figure 2.3.)

So that you can find reasons for (or against) believing what you read or hear (or think)! And why is it important to have these reasons? For one thing, it can make you feel more confident about your beliefs and the beliefs of others. For another, it can help you try to convince others about your beliefs—not necessarily to convince them that they should believe what you believe, but to help them understand why you believe what you do.

I do not pretend that I can refute these two views; but I can challenge them . . . . (Popper, 1978, §4, p. 148)

This is the heart of philosophy: not (necessarily) coming up with answers, but challenging assumptions and forcing you to think about alternatives. My father’s favorite admonition was: Never make assumptions. That is, never assume that something is the case or that someone is going to do something; rather, try to find out if it is the case, or ask the person. In other words, challenge all assumptions. Philosophers, as James Baldwin (1962) said about artists, “cannot and must not take anything for granted, but must drive to the heart of every answer and expose the question the answer hides.”

This is one way that progress can be made in philosophy: It may be backward-looking progress, because, instead of looking “forward” to implications of your assumptions, you look “backward” to see where those assumptions might have come from.

Besides these two directions of progress, there can be a third, which is orthogonal to these two: “Sideway” progress can be made by considering other issues that might
not underlie (“backward”) or follow from (“forward”) the one that you are considering, but that are “inspired” or “suggested” by it.

2.6 What Is “Rational”?

Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and the further conclusions to which it tends, constitutes reflective thought.

—John Dewey (1910, p. 6)

Mere statements (that is, opinions) by themselves are not rational. Rather, arguments—reasoned or supported statements—are capable of being rational. As the American philosopher John Dewey suggested, it’s not enough to merely think something; you must also consider reasons for believing it (looking “backward”), and you must also consider the consequences of believing it (looking “forward”). That is, being rational requires logic.

But there are lots of different (kinds of) logics, so there are lots of different kinds of rationality. And there is another kind of rationality, which depends on logics of various kinds, but goes beyond them in at least one way: empirical, or scientific, rationality. Let’s look at these two kinds of rationality.

2.6.1 Kinds of Rationality

Philosophy: the ungainly attempt to tackle questions that come naturally to children, using methods that come naturally to lawyers.


There are (at least) two basic kinds of logical rationality: deductive (or absolutely certain) rationality and scientific (or probabilistic) rationality. There is also, I think, a third kind, which I’ll call “psychological” or maybe “economic”, and which is at the heart of knowledge representation and reasoning in AI.

2.6.1.1 Deductive Rationality

“Deductive” logic is the main kind of logical rationality. Reasons $P_1, \ldots, P_n$ deductively support (or “yield”, or “entail”, or “imply”) a conclusion $C$ iff $C$ must be true if all of the $P_i$ are true. The technical term for this is ‘validity’: A deductive argument is said to be valid iff it is impossible for the conclusion to be false while all of the premises are true. This can be said in a variety of ways: A deductive argument is valid iff, whenever all of its premises are true, its conclusion cannot be false. Or: A deductive argument is valid iff, whenever all of its premises are true, its conclusion must also be true. Or: A deductive argument is valid iff the rules of inference that lead from its premises to its conclusion preserve truth.

For example, the rule of inference called “Modus Ponens” says that, from $P$ and ‘if $P$, then $C$‘, you may deductively infer $C$. Using the symbol ‘$\vdash$’ to represent this
truth-preserving relation between reasons (usually called ‘premises’) and a conclusion that is deductively supported by them, the logical notation for Modus Ponens is:

\[ P, \ (P \rightarrow C) \vdash_D C \]

For example, let \( P = \text{“Today is Wednesday.”} \) and let \( C = \text{“We are studying philosophy.”} \) so the inference becomes: “Today is Wednesday. If today is Wednesday, then we are studying philosophy. Therefore (deductively), we are studying philosophy.” (For more on Modus Ponens, see §2.10.4.)

There are three somewhat surprising things about validity (or deductive rationality) that must be pointed out:

1. **Any or all of the premises \( P_i \) of a valid argument can be false!** In the second version of the characterization of validity above, note that the conditional term ‘whenever’ allows for the possibility that one or more premises are false. So, any or all of the premises of a deductively valid argument can be false, as long as, if they were to be true, then the conclusion would also have to be true.

2. **The conclusion \( C \) of a valid argument can be false!** How can a “truth-preserving” rule lead to a false conclusion? By the principal familiar to computer programmers known as “garbage in, garbage out”: *If one of the \( P_i \) is false, even truth-preserving rules of inference can lead to a false \( C \).*

   As is the case with any sentence, the conclusion of an argument can, of course, be true or false, (or, more leniently, you can agree with it or not). But, besides being “absolutely” or “independently” true or false (or agreeable or disagreeable), a conclusion can also be relatively true. More precisely: a conclusion can be true relative to the truth of its premises. What this means is that you can have a situation in which a sentence is, let’s say, “absolutely” or “independently” false (or you disagree with it), but it could also be true relative to some premises.

   How could that be? Easy: If the world is such that, whenever it makes the premises true, then it also makes the conclusion true, then the conclusion is true relative to the premises. But note that this is a conditional statement: “Whenever the world makes the premises true, then . . .”. The premises provide a background “context” in which to evaluate the conclusion. The conclusion \( C \) only has to be true relative to the premises (that is, true relative to its context). In other words, \( C \) would be true if all of the \( P_i \) were true. But sometimes the world might not make the premises true. And then we can’t say anything about the truth of the conclusion. When a conclusion is true relative to its premises, then the argument is said to be valid.

   So, when can we be sure that the conclusion \( C \) of a valid argument is really true (and not just “relatively” true)? The answer is that \( C \) is true iff (1) all of the \( P_i \) are true, and (2) the rules of inference that lead from the \( P_i \) to \( C \) “preserve” truth. Such a deductive argument is said to be “sound”, that is, it is valid and all of its premises are, in fact, true.

3. **The premises \( P_i \) of a valid argument can be irrelevant to the conclusion \( C \)!** But that’s not a good idea, because it wouldn’t be a convincing argument. The
classic example of this is that anything follows deductively from a contradiction: From the two contradictory propositions ‘2 + 2 = 4’ and ‘2 + 2 ≠ 4’, it can be deductively inferred that the philosopher Bertrand Russell (a noted atheist) is the Pope.

Proof and Further Reading:
Let $P$ and $¬P$ be the two premises, and let $C$ be the conclusion. From $P$, we can deductively infer $(P ∨ C)$, by the truth-preserving rule of Addition (a form of $∨$-introduction). Then, from $(P ∨ C)$ and $¬P$, we can deductively infer $C$, by the truth-preserving rule of Disjunctive Syllogism (a form of $∨$-elimination). So, in the “Pope Russell” argument, from ‘2 + 2 = 4’, we can infer that either $2 + 2 = 4$ or Russell is the Pope (or both). That is, we can infer that at least one of those two propositions is true. But we have also assumed that one of them is false: $2 + 2 ≠ 4$. So it must be the other one that is true: Therefore, Russell must be the Pope! (But remember point 2, above: It doesn’t follow from this argument that Russell is the Pope. All that follows is that Russell would be the Pope (and so would you!) if $2 + 2$ both does and does not equal 4.)

“Relevance” logics are one way of dealing with this problem; see Anderson and Belnap 1975; Anderson et al. 1992. For applications of relevance logic to AI, see Shapiro and Wand 1976; Martins and Shapiro 1988.

We’ll say a lot more about this in the Appendix to this chapter (§2.10).

2.6.1.2 Inductive Logical Rationality

“Inductive” logic is one of the two main kinds of scientific rationality. The other is “abductive” logic (to be discussed in the next section). Deductive rationality, which is more characteristic of mathematics than of the experimental sciences, is, however, certainly part of science.

In inductive logic, $P_1, . . . , P_n ⊨_I C$ iff $C$ is probably true if all of the $P_i$ are true. For example, suppose that you have an urn containing over a million ping-pong balls, and suppose that you remove one of them at random and observe that it is red. What do you think the chances are that the next ball will also be red? They are probably not very high. But suppose that the second ball that you examine is also red. And the third. . . And the 999,999th. Now how likely do you think it is that the next ball will also be red? The chances are probably very high, so:

$\text{Red(ball}_1), . . . , \text{Red(ball}_{999,999}) ⊨_I \text{Red(ball}_{1,000,000})$.

Unlike deductive inferences, however, inductive ones do not guarantee the truth of their conclusion. Although it is not likely, it is quite possible that the millionth ping-pong ball will be, say, the only blue one in the urn.
2.6.1.3 Abductive Logical Rationality

Adding a new hypothesis or axiom to a theory for the purpose of explaining already known facts is a process known as “abduction”.

—Aaron Sloman (2010, slide 56)

“Abductive” logic, sometimes also known as “inference to the best explanation”, is also scientific: From observation \( O \) made at time \( t_1 \), and from a theory \( T \) that deductively or inductively entails \( O \), one can abductively infer that \( T \) must have been the case at earlier time \( t_0 \). In other words, \( T \) is an explanation of why you have observed \( O \). Of course, it is not necessarily a good, much less the best, explanation, but the more observations that \( T \) explains, the better a theory it is. (But what is a “theory”? We’ll delve into that in §4.7. For now, you can think of a theory as just a set of statements that describe, explain, or predict some phenomenon.)

Abductive arguments are deductively invalid; they have the form (A):

\[
(A) \quad O, \; (T \rightarrow O) \not\vdash_D \; T
\]

Argument (A) is called the fallacy of affirming the consequent.

Digression on Affirming the Consequent:

\( O \) is the “consequent” of the conditional statement \( (T \rightarrow O) \). “Affirming” \( O \) as a premise thus “affirms the consequent”. (We will come back to this in §4.9.1.1.) But if \( O \) is true and \( T \) is false, then both premises are true, yet the conclusion \( (T) \) is not.

In another form of abduction, from observation \( O_1 \) made at time \( t_1 \), and from observation \( O_2 \) made at a later time \( t_2 \), one can abductively infer that \( O_1 \) might have caused or logically entailed \( O_2 \). This, too, is deductively invalid: Just because two observations are correlated does not imply that the first causes the second, because the second might have caused the first, or both might have been caused by a third thing.

Like inductive inferences, abductive ones are not deductively valid and do not guarantee the truth of their conclusion. But abductive inferences are at the heart of the scientific method for developing and confirming theories. And they are used in the law, where they are known as “circumstantial evidence”.

Further Reading:

For the origin of the term in the writings of the American philosopher Charles Sanders Peirce (who pronounced his name like the word ‘purse’), see http://www.helsinki.fi/science/commens/terms/abduction.html. For more on abductive logic, see Harman 1965; Lipton 2004; Campos 2011.

2.6.1.4 Non-Monotonic Logical Rationality

“Non-monotonic” reasoning is more “psychologically real” than any of the others. It also underlies what the economist and AI researcher Herbert Simon called “satisficing” (or being satisfied with something that suffices to answer your question rather than having an optimal answer), for which he won the Nobel Prize in Economics.
2.6. WHAT IS “RATIONAL”?

In monotonic logics (such as deductive logics), once you have proven that a conclusion $C$ follows from a premise $P$, then you can be assured that it will always so follow. But in non-monotonic logic, you might infer conclusion $C$ from premise $P$ at time $t_0$, but, at later time $t_1$, you might learn that it is not the case that $C$. In that case, you must revise your beliefs. For example, you might believe that birds fly and that Tweety is a bird, from which you might conclude that Tweety flies. But if you then learn that Tweety is a penguin, you will need to revise your beliefs.

Further Reading:
For a history of satisficing, see Brown 2004. We'll return to this topic in §§3.15.2.3, 5.7, and 11.4.5.2. A great deal of work on non-monotonic logics has been done by researchers in the branch of AI called “knowledge representation”; see the bibliography at http://www.cse.buffalo.edu/~rapaport/663/F08/nonmono.html

2.6.1.5 Computational Rationality

In addition to logical rationality and scientific rationality, the astronomer Kevin Heng argues that,

a third, modern way of testing and establishing scientific truth—in addition to theory and experiment—is via simulations, the use of (often large) computers to mimic nature. It is a synthetic universe in a computer. . . . If all of the relevant physical laws are faithfully captured [in the computer program] then one ends up with an emulation—a perfect, *The Matrix*-like replication of the physical world in virtual reality. (Heng, 2014, p. 174)

One consideration that this raises is whether this is really a third way, or just a version of logical rationality, perhaps extended to include computation as a kind of “logic”. (We’ll discuss computer programs and computational simulations in Chapter 15, and we’ll return to *The Matrix* in §20.8.)

However, all of the above kinds of rationality seem to have one thing in common: They are all “declarative”. That is, they are all concerned with statements (or propositions) that are true or false. But the philosopher Gilbert Ryle (1945, especially p. 9) has argued that there is another kind of rationality, one that is “procedural” in nature: It has been summarized as “knowing how” (to do something), rather than “knowing that” (something is the case). We will explore this kind of rationality in more detail in §§3.6.1 and 3.14.4.

2.6.2 Science and Philosophy

If philosophy is a search for truth by rational means, what is the difference between philosophy and science? After all, science is also a search for truth by rational means! Is philosophy worth doing? Or can science answer all of our questions?
2.6.2.1 Is Science Philosophy?

Is the experimental or empirical methodology of science “rational”? It is not (entirely) deductive. But it yields highly likely conclusions, and is often the best we can get.

I would say that science is philosophy, as long as experiments and empirical methods are considered to be “rational” and yield truth. Physics and psychology, in fact, used to be branches of philosophy: Isaac Newton’s *Principia*—the book that founded modern physics—was subtitled “Mathematical Principles of Natural Philosophy” (italics added), not “Mathematical Principles of Physics”, and psychology split off from philosophy only at the turn of the 20th century. The philosophers Aristotle (384–322 BCE, around 2400 years ago) and Kant (1724–1804, around 250 years ago) wrote physics books. The physicists Einstein and Mach wrote philosophy. And the “philosophy naturalized” movement in contemporary philosophy (championed by the philosopher Willard Van Orman Quine) sees philosophy as being on a continuum with science. (See §2.6.2.2; we’ll come back to this in §2.8.)

But, if experiments don’t count as being rational, and only logic counts, then science is not philosophy. And science is also not philosophy, if philosophy is considered to be the search for universal or necessary truths, that is, things that would be true no matter what results science came up with or what fundamental assumptions we made.

There might be conflicting world views (for example, creationism vs. evolution, perhaps). Therefore, the best theory is one that is (1) consistent, (2) as complete as possible (that is, that explains as much as possible), and (3) best-supported by good evidence.

You can’t refute a theory. You can only point out problems with it and then offer a better theory. Suppose that you infer a prediction $P$ from a theory $T$ together with a hypothesis $H$, and then suppose that $P$ doesn’t come true (your experiment fails; that is, the experimental evidence is that $P$ is not the case). Then, logically, either $H$ is not the case or $T$ is not the case (or both!). And, since $T$ is probably a complex conjunction of claims $A_1, \ldots, A_n$, then, if $T$ is not the case, then at least one of the $A_i$ is not the case. In other words, you need not give up a theory; you only need to revise it. That is, if $P$ has been falsified, then you only need to give up one of the $A_i$ or $H$, not necessarily the whole theory $T$.

However, sometimes you should give up an entire theory. This is what happens in the case of “scientific revolutions”, such as (most famously) when Copernicus’s theory that the Earth revolves around the Sun (and not vice versa) replaced the Ptolemaic theory, small revisions to which were making it overly complex without significantly improving it. (We’ll say more about this in §4.9.2.)

2.6.2.2 Is Philosophy a Science?

Could philosophy be more scientific (that is, experimental) than it is? Should it be? The philosopher Colin McGinn (2012a) takes philosophy to be a science (“a systematically organized body of knowledge”), in particular, what he dubs ‘ontical science’: “the subject consists of the search for the essences of things by means of a priori methods” (McGinn, 2012b). In a later paper, he argues that philosophy is a science just like physics or mathematics. More precisely, he says that it is the logical science of concepts

There is a relatively recent movement (with some older antecedents) to have philosophers do scientific (mostly psychological) experiments in order to find out, among other things, what “ordinary” people (for example, people who are not professional philosophers) believe about certain philosophical topics.

Further Reading:
For more information on this movement, sometimes called ‘X-Phi’, see Nahmias et al. 2006; Appiah 2007, 2008; Knobe 2009; Beebe 2011; Nichols 2011; Roberts and Knobe 2016. For an argument against experimental philosophy, see Deutsch 2009. Whether or not X-Phi is really philosophy, it is certainly an interesting and valuable branch of cognitive science.

But there is another way that philosophy can be part of a scientific worldview. This can be done by philosophy being continuous with science, that is, by being aware of, and making philosophical use of, scientific results. Rather than being a passive, “armchair” discipline that merely analyzes what others say and do, philosophy can—and probably should—be a more active discipline, even helping to contribute to science (and other disciplines that it thinks about).

Further Reading:
For a useful discussion of this, which is sometimes called “naturalistic philosophy”, see Thagard 2012. Williamson (2007) argues that there’s nothing wrong with “armchair” philosophy.

Philosophers can also be more “practical” in the public sphere: “The philosophers have only interpreted the world in various ways; the point is to change it” (Marx, 1845). But an opposing point of view considers that “philosophers . . . are ordained as priests to keep alive the sacred fires in the altar of impartial truth” (“Philonous”, 1919, p. 19)! (For more on this, see §5.7.)

Further Reading:
For a debate on science vs. philosophy, read Linker 2014; Powell 2014; Pigliucci 2014, in that order. For a discussion of whether philosophy or science is “harder”, see Papineau 2017.

2.6.3 Is It Always Rational to Be Rational?

Is there anything to be said in favor of not being rational?

Suppose that you are having trouble deciding between two apparently equal choices. This is similar to a problem from mediaeval philosophy known as “Buridan’s Ass” (see Zupko 2011): According to one version, an ass (that is, a donkey) was placed equidistant between two equally tempting bales of hay but died of starvation because it couldn’t decide between the two of them. My favorite way out of such a quandary is to imagine tossing a coin and seeing how you feel about how it lands: If it lands heads up, say, but you get a sinking feeling when you see that, because you would rather that it had landed tails up, then you know what you would have preferred, even if you had “rationally” decided that both choices were perfectly equally balanced.
Further Reading:

2.7 What Is the Import of “Personal Search”?  

... I’m not trying to change anyone’s mind on this question. I gave that up long ago. I’m simply trying to say what I think is true.
—Galen Strawson (2012, p. 146)

And among the philosophers, there are too many Platos to enumerate. All that I can do is try to give you mine.
—Rebecca Newberger Goldstein (2014, p. 396)

[My] purpose is to put my own intellectual home in order ....
—Hilary Putnam (2015)

“The philosophy of every thinker is the more or less unconscious autobiography of its author,” Nietzsche observed .... —Clancy Martin (2015)

The philosopher Hector-Neri Castañeda used to say that philosophy should be done “in the first person, for the first person” (Rapaport, 2005a). So, philosophy is whatever I am interested in, as long as I study it in a rational manner and aim at truth (or, at least, aim at the best theory).

There is another way in which philosophy must be a personal search for truth. As one introductory book puts it, “the object here is not to give answers ... but to introduce you to the problems in a very preliminary way so that you can worry about them yourself” (Nagel, 1987, pp. 6–7, my italics). The point is not to hope that someone else will tell you the answers to your questions. That would be nice, of course; but why should you believe them? The point, rather, is for you to figure out answers for yourself.

It may be objected that your first-person view on some topic, no matter how well thought out, is, after all, just your view. “Such an analysis can be of only parochial interest” (Strevens, 2019) or might be seriously misleading (Dennett, 2017, pp. 364–370). Another philosopher, Hilary Kornblith, agrees:

I believe that the first-person perspective is just one perspective among many, and it is wholly undeserving of the special place which these philosophers would give it. More than this, this perspective is one which fundamentally distorts our view of crucial features of our epistemic situation. Far from lauding the first-person perspective, we should seek to overcome its defects. (Kornblith, 2013, p. 126)

But there is another important feature of philosophy, as I mentioned in §1.3: It is a conversation. And if you want to contribute to that conversation, you will have to take others’ views into account, and you will have to allow others to make you think harder about your own views.
2.8. WHAT IS THE IMPORT OF “IN ANY FIELD”?

The desire for an “Authority” to answer all questions for you has been called the “Dualistic” stance towards knowledge. But the Dualist soon realizes that not all questions have answers that everyone agrees with, and some questions don’t seem to have answers at all (at least, not yet).

Rather than stagnating in a middle stance of “Multiplism” (a position that says that, because not all questions have answers, multiple opinions—proposed answers—are all equally good), a further stance is that of “Contextual Relativism”: All proposed answers or opinions can (should!) be considered—and evaluated!—relative to and in the context of assumptions, reasons, or evidence that can support them.

Eventually, you “Commit” to one of these answers, and you become responsible for defending your commitment against “Challenges”. But that is (just) more thinking and analysis—more philosophizing. Moreover, the commitment that you make is a personal one (one that you are responsible for). As the computer scientist Richard W. Hamming warned, “In science and mathematics we do not appeal to authority, but rather you are responsible for what you believe” (Hamming, 1998, p. 650).

Further Reading:
The double-quoted and capitalized terms come from William Perry (see §2.5, above). For more on Perry’s theory, see Perry 1970, 1981; §C, below; and http://www.cse.buffalo.edu/~rapaport/perry.positions.html. See also the answer to a question about deciding which of your own opinions to really believe, at http://www.askphilosophers.org/question/5563.

It is in this way that philosophy is done “in the first person, for the first person”, as Castañeda said.

2.8. What Is the Import of “In Any Field”?

One of the things about philosophy is that you don’t have to give up on any other field. Whatever field there is, there’s a corresponding field of philosophy. Philosophy of language, philosophy of politics, philosophy of math. All the things I wanted to know about I could still study within a philosophical framework.
—Rebecca Newberger Goldstein, cited in Reese 2014b

[He] is a philosopher, so he’s interested in everything . . . .
—David Chalmers (describing the philosopher Andy Clark), as cited in Cane 2014.

It is not really possible to regret being a philosopher if you have a theoretical (rather than practical or experiential) orientation to the world, because there are no boundaries to the theoretical scope of philosophy. For all X, there is a philosophy of X, which involves the theoretical investigation into the nature of X. There is philosophy of mind, philosophy of literature, of sport, of race, of ethics, of mathematics, of science in general, of specific sciences such as physics, chemistry and biology; there is logic and ethics and aesthetics and philosophy of history and history of philosophy. I can read Plato and Aristotle and Galileo and Newton and Leibniz and Darwin and Einstein and John Bell and just be doing my job. I could get fed up with all that and read Eco and Foucault and Aristophanes and Shakespeare for
Philosophy also studies things that are not studied by any single discipline; these are sometimes called “the Big Questions”: What is truth? What is beauty? What is good (or just, or moral, or right)? What is the meaning of life? What is the nature of mind? (For a humorous take on this, see Fig. 2.4.) Or, as the philosopher Jim Holt put it: “Broadly speaking, philosophy has three concerns: how the world hangs together, how our beliefs can be justified, and how to live” (Holt, 2009). The first of these is metaphysics, the second is epistemology, and the third is ethics. (Similar remarks have been made by Flanagan 2012, p. B4; Schwitzgebel 2012; Weatherson 2012.)

But the main branches of philosophy go beyond these “big three”:

1. **Metaphysics** tries to “understand the nature of reality in the broadest sense: what kinds of things and facts ultimately constitute everything there is” (Nagel, 2016, p. 77). It tries to answer the question “What is there?” (and also the question “Why is there anything at all?”). Some of the things that there might be include: physical objects, properties, relations, individuals, time, God, actions, events, minds, bodies, etc. There are major philosophical issues surrounding each of these. Here are just a few examples:

   - Which physical objects “really” exist? Do rocks and people exist? Or are they merely collections of molecules? But molecules are constituted by atoms; and atoms by electrons, protons, and neutrons. And, according to the “standard model”, the only really elementary particles are quarks, leptons (which include electrons), and gauge bosons; so maybe those are the only really existing physical objects. Here is a computationally relevant version of this kind of question: Do computer programs that deal with, say, student records model students? Or are they just dealing with 0s and 1s? (We’ll discuss this in §14.3.3.) And, on perhaps a more fanciful level, could a computer program model students so well that the “virtual” students in the program believe that they are real? (If this sounds like the film *The Matrix*, see §20.8.)
   - Do “socially constructed” things like money, universities, governments,
2.8. WHAT IS THE IMPORT OF “IN ANY FIELD”?

etc., really exist (in the same way that people or rocks do)? (This problem is discussed in Searle 1995.)

- Do properties really exist? Or are they just collections of similar (physical) objects. In other words, is there a property—“Redness”—in addition to the class of individual red things? Sometimes, this is expressed as the problem of whether properties are “intensional” (like Redness) or “extensional” (like the set of individual red things). (See §3.4 for more about this distinction.)

- Are there any important differences between “accidental” properties (such as my property of being a professor of computer science rather than my being a professor of philosophy) and “essential” properties (such as my property of being a human rather than being a laurel tree)?

- Do “non-existents” (such as Santa Claus, unicorns, Sherlock Holmes, etc.) exist in some sense? After all, we can and do think and talk about them. Therefore, whether or not they “exist” in the real world, they do need to be dealt with.

- **Ontology** is the branch of metaphysics that is concerned with the objects and kinds of objects that exist according to one’s metaphysical (or even physical) theory, their properties, and their relations to each other (such as whether some of them are “sub-kinds” of others, inheriting their properties and relations from their “super-kinds”). For example, the modern ontology of physics recognizes the existence only of fermions (quarks, leptons, etc.) and bosons (photons, gluons, etc.); everything else is composed of things (like atoms) that are, in turn, composed of these. Ontology is studied both by philosophers and by computer scientists. In software engineering, “object-oriented” programming languages are more focused on the kinds of objects that a program must deal with than with the instructions that describe their behavior. In AI, ontology is a branch of knowledge representation that tries to categorize the objects that a knowledge-representation theory is concerned with.

Further Reading:

For a computational approach to the question “What is there?”, see http://www.cse.buffalo.edu/~rapaport/663/F06/course-summary.html. For an interesting take on what “really” exists, see Unger 1979a,b. On non-existence, see Quine 1948. For a survey of the AI approach to non-existence, see Hirst 1991. And for some papers on a fully intensional AI approach to these issues, see Maida and Shapiro 1982; Rapaport 1986a; Wiebe and Rapaport 1986; Shapiro and Rapaport 1987, 1991; Rapaport et al. 1997. For more information on ontology, see http://www.cse.buffalo.edu/~rapaport/563S05/ontology.html. For the AI version of ontology, see http://aitopics.org/topic/ontologies and http://ontology.buffalo.edu/.

And so on. As William James said:

---

4 http://www.theoi.com/Nymphe/NympheDaphne.html

5 https://en.wikipedia.org/wiki/Elementary_particle
Metaphysics means only an unusually obstinate attempt to think clearly and consistently. ... A geologist’s purposes fall short of understanding Time itself. A mechanist need not know how action and reaction are possible at all. A psychologist has enough to do without asking how both he [sic] and the mind which he studies are able to take cognizance of the same outer world. But it is obvious that problems irrelevant from one standpoint may be essential for another. And as soon as one’s purpose is the attainment of the maximum of possible insight into the world as a whole, the metaphysical puzzles become the most urgent ones of all. (James, 1892, “Epilogue: Psychology and Philosophy”, p. 427; my italics)

2. **Epistemology** is the study of knowledge and belief:

   Epistemology is concerned with the question of how, since we live, so to speak, inside our heads, we acquire knowledge of what there is outside our heads. (Simon, 1996a, p. 162)

   How do we know what there is? How do we know that there is anything? What is knowledge? Is it justified, true belief (as Plato thought)? Or are there counterexamples to that analysis? That is, can you be logically justified in believing something that is in fact true, and yet not know it? (See Gettier 1963.) Are there other kinds of knowledge, such as knowing how to do something (see §3.14.4), knowing a person by acquaintance, or knowing who someone is by description? What is belief, and how does it relate to knowledge? Can a computer (or a robot) be said to have beliefs or knowledge? In fact, the branch of AI called “knowledge representation” applies philosophical results in epistemology to issues in AI and computer science in general, and it has contributed many results to philosophy as well.

Further Reading:

On knowledge representation, see Buchanan 2006; Shoham 2016; and the bibliography at http://www.cse.buffalo.edu/~rapaport/663/F08/krresources.html.

3. **Ethics** tries to answer “What is good?”, “What ought we to do?”. We'll look at some ethical issues arising from computer science in Chapters 18 and 20.

4. Ethics is closely related to both social and political philosophy and to the philosophy of law, which try to answer “What are societies?”, “What are the relations between societies and the individuals who constitute them?”, “What is the nature of law?”.

5. **Aesthetics** (or the philosophy of art) tries to answer “What is beauty?”, “What is art?”. (On whether computer programs, like mathematical theorems or proofs, can be “beautiful”, see §3.14.2.)

6. **Logic** is the study of good reasoning: What is truth? What is rationality? Which arguments are good ones? Can logic be computationally automated? (Recall our discussion in §2.6.)
7. Philosophy is one of the few disciplines (history is another) in which the history of itself is one of its branches: the history of philosophy looks at what famous philosophers of the past believed, and tries to reinterpret their views in the light of contemporary thinking.

8. And of central interest for the philosophy of computer science, there are numerous “philosophies of”:

- **Philosophy of language** tries to answer “What is language?”, “What is meaning?”. It has large overlaps with linguistics and with cognitive science (including AI and computational linguistics).

- **Philosophy of mathematics** tries to answer “What is mathematics?”, “Is math about numbers, numerals, sets, structures?”, “What are numbers?”, “Why is mathematics so applicable to the real world?”.

Further Reading:
On the philosophy of mathematics, see Benacerraf and Putnam 1984; Pincock 2011; Horsten 2015.

- **Philosophy of mind** tries to answer “What is the mind related to the brain?” (this is known as the “mind-body” problem). Are minds and bodies two different kinds of substances? (This is known as “dualism”, initially made famous by Descartes.) Or are they two different aspects of some one, underlying substance? (This is a position made famous by the 17th-century Dutch philosopher Baruch Spinoza.) Or are there no minds at all, but only brains? (This is known as “materialism” or “physicalism”; it is the position of most contemporary philosophers and scientists.) Or are there no independently existing physical objects, but only ideas in our minds? (This is known as “idealism”, made famous by the 18th-century Irish philosopher George Berkeley.) (In §12.4.6, we’ll say more about the mind-body problem and its relation to the software-hardware distinction.) The philosophy of mind also investigates whether computers can think (or be said to think), and it has close ties with cognitive science and AI, issues that we will take up in Chapter 19.

- **Philosophy of science** tries to answer “What is science?”, “What is a scientific theory?”, “What is a scientific explanation?”. The philosophy of computer science is part of the philosophy of science. The philosopher Daniel C. Dennett has written that there was a “reform that turned philosophy of science from an armchair fantasy field into a serious partnership with actual science. There came a time when philosophers of science decided that they really had to know a lot of current science from the inside” (Dennett, 2012, p. 12). Although you do not need to know a lot about computer science (or philosophy, for that matter) to learn something from the present book, clearly the more you know about each topic, the more you will be able both to understand what others are saying and to contribute to the conversation. (We will look at the philosophy of science in Chapter 4.)
In general, for any $X$, there can be a philosophy of $X$: the philosophical investigation of the fundamental assumptions, methods, and goals of $X$ (including metaphysical, epistemological, and ethical issues), where $X$ could be: biology, education, history, law, physics, psychology, religion, etc., including, of course, AI and computer science. The possibility of a philosophy of $X$ for any $X$ is the main reason why philosophy is the rational search for truth in any field. “Philosophy is 99 per cent about critical reflection on anything you care to be interested in” (Richard Bradley, in Popova 2012). Philosophy in general, and especially the philosophy of $X$, is a “meta-discipline”: In a discipline $X$, you think about $X$ (in the discipline of mathematics, you think about mathematics); but in the philosophy of $X$, you think about thinking about $X$. Even those subjects that might be purely philosophical (metaphysics, epistemology, and ethics) have strong links to disciplines like physics, psychology, and political science, among others.

$X$, by the way, could also be . . . philosophy! The philosophy of philosophy, also known as “metaphilosophy”, is exemplified by this very chapter, which is an investigation into what philosophy is and how it can be done. Some people might think that the philosophy of philosophy is the height of “gazing at your navel”, but it’s really what’s involved when you think about thinking, and, after all, isn’t AI just computational thinking about thinking?

Philosophy, besides being interested in any specific topic, also has an overarching or topic-spanning function: It asks questions that don’t fall under the aegis of specific topics and that span multiple topics: The philosopher Wilfrid Sellars said, “The aim of philosophy, abstractly formulated, is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term” (Sellars, 1963, p. 1). So, for instance, while it is primarily (but not only) mathematicians who are interested in mathematics per se and primarily (but not only) scientists who are interested in science per se, it is primarily (but not only) philosophers who are interested in how and why mathematics is so useful for science (see P. Smith 2010).

Are there any topics that philosophy doesn’t touch on? I’m sure that there are some topics that philosophy hasn’t touched on. But I’m equally sure that there are no topics that philosophy couldn’t touch on.
Further Reading:

Russell 1946 explains why studying philosophy is important for everyone, not just professional philosophers. McGinn 2003 is a brief autobiography of how a well-known contemporary philosopher got into the field.

The website AskPhilosophers (http://www.askphilosophers.org/) has suggested answers to some relevant questions:

1. What do people mean when they speak of “doing” philosophy?,
   http://www.askphilosophers.org/question/2915

2. Why are philosophers so dodgy when asked a question?
   http://www.askphilosophers.org/question/2941

3. Are there false or illegitimate philosophies, and if so, who’s to say which ones are valid and which are invalid? http://www.askphilosophers.org/question/2994

4. What does it take to be a philosopher? http://www.askphilosophers.org/question/4609

2.9 Philosophy and Computer Science

[If there remain any philosophers who are not familiar with some of the main developments in artificial intelligence, it will be fair to accuse them of professional incompetence, and that to teach courses in philosophy of mind, epistemology, aesthetics, philosophy of science, philosophy of language, ethics, metaphysics, and other main areas of philosophy, without discussing the relevant aspects of artificial intelligence will be as irresponsible as giving a degree course in physics which includes no quantum theory.

—Aaron Sloman (1978, §1.2, p. 3)

Philosophy and computer science overlap not only in some topics of common interest (logic, philosophy of mind, philosophy of language, etc.), but also in methodology: the ability to find counterexamples; refining problems into smaller, more manageable ones; seeing implications; methods of formal logic; and so on.

For example here’s an application of predicate logic to artificial intelligence (AI): In the late 1950s, one of the founders of AI, John McCarthy, proposed a computer program to be called “The Advice Taker”, as part of a project that he called “programs with common sense”. The idea behind The Advice Taker was that problems to be solved would be expressed in a predicate-logic language (only a little bit more expressive than first-order logic), a set of premises or assumptions describing required background information would be given, and then the problem would be solved by logically deducing an answer from the assumptions.
He gave an example: getting from his desk at home to the airport. It begins with premises like
\[ \text{at}(I, \text{desk}) \]
meaning “I am at my desk”, and rules like
\[ \forall x \forall y \forall z [\text{at}(x, y) \land \text{at}(y, z) \rightarrow \text{at}(x, z)], \]
which expresses the transitivity of the “at” predicate (for any three things \(x, y,\) and \(z\), if \(x\) is at \(y\), and \(y\) is at \(z\), then \(x\) is at \(z\)), along with slightly more complicated rules (which go slightly beyond the expressive power of first-order logic) such as:
\[ \forall x \forall y \forall z [\text{walkable}(x) \land \text{at}(y, x) \land \text{at}(z, x) \land \text{at}(I, y) \rightarrow \text{can}(I, \text{go}(y, z, \text{walking}))] \]
(that is, if \(x\) is walkable, and if \(y\) and \(z\) are at \(x\), and if \(I\) am at \(y\), then \(I\) can go from \(y\) to \(z\) by walking).

The proposition to be proved from these (plus lots of others) is:
\[ \text{want}(\text{at}(I, \text{airport})) \]
(that is, we want it to be the case that \(I\) am at the airport).

Further Reading:
To see how to get to the airport, take a look at McCarthy 1959. McCarthy is famous for at least the following things: He came up with the name ‘artificial intelligence’, he invented the programming language Lisp, and he helped develop time sharing. For more information on him, see http://en.wikipedia.org/wiki/John_McCarthy_(computer_scientist) and http://aitopics.org/search/site/John%20McCarthy.

I have mentioned a few different kinds of logic: Propositional logic is the logic of sentences, treating them “atomically” as simply being either true or false, and as not having any “parts”. First-order predicate logic can be thought of as a kind of “sub-atomic” logic, treating sentences as being composed of terms standing in relations. But there are also second-order logics, modal logics, relevance logics, and many more (not to mention varieties of each). Is one of them the “right” logic? Tharp 1975 asks that question, which can be expressed as a “thesis” analogous to the Church-Turing Computability Thesis: Where the Computability Thesis asks if the formal theory of Turing Machine computability entirely captures the informal, pre-theoretic notion of computability, Tharp asks if there is a formal logic that entirely captures the informal, pre-theoretic notion of logic. We’ll return to some of these issues in Chapter 11.

For further discussion of the value of philosophy for computer science (and vice versa!), see Arner and Slein 1984, especially pp. 76–77.

In the next chapter, we’ll begin our philosophical investigation into computer science.
A Philosophical Round

I sat upon a chair . . .

(but was it there?
and what is ‘I’?
and is ‘I’ me?)
. . . and had some thoughts on
PHILOSOPHY
(where ‘had’ means ‘do’?
and ‘thoughts’: insights, or recall?
and the ‘Big P’ too:
defined by others, or by me?
or some view
overall?)
And I wondered:
Is it always best when plainly told? . . .
(but best for what? for whom?
and ‘it’ means all, or some?
‘plainly’ means clear, or dry?
‘told’ means typed? orated?
how confidently stated?
and who should have this say?)
. . . Or have fictional forms a part to play?
(but ‘fiction’: poetry? theatre?
music? art? prose?
comedy? tragedy? adventure?
long? short? episodic?
concise? verbose?
literal, or metaphoric?
epistolic? dialectic? parabolic . . . ?)
WAIT!
This has become more abstruse than Zen.
I think I’d better start again:
I sat upon a chair . . .

—Daryn Green (2014a)
2.10 Appendix: Argument Analysis and Evaluation

2.10.1 Introduction

In §2.3, I said that the methodology of philosophy involved “rational means” for seeking truth, and in §2.6, we looked at different kinds of rational methods. In this appendix, we’ll look more closely at one of those methods—argument analysis and evaluation—which you will be able to practice when you do the exercises in Appendix A. Perhaps more importantly for some readers, argument analysis is a topic in two of the knowledge areas (Discrete Structures, and Social Issues and Professional Practice/Analytical Tools) of Computer Science Curricula 2013 (https://ieeecs-media.computer.org/assets/pdf/CS2013-final-report.pdf).

Unless you are gullible—willing to believe everything you read or anything that an authority figure tells you—you should want to know why you should believe something that you read or something that you are told. If someone tells you that some proposition $C$ is true because some other propositions $P_1$ and $P_2$ are true, you might then consider, first, whether those reasons ($P_1$ and $P_2$) really do support the conclusion $C$ and, second, whether you believe the reasons.

Let’s consider how you might go about doing this.

2.10.2 A Question-Answer Game

Consider two players, $Q$ and $A$, in a question-answer game:

**Step 1** $Q$ asks whether $C$ is true.

**Step 2** $A$ responds: “$C$, because $P_1$ and $P_2$.”
2.10. APPENDIX: ARGUMENT ANALYSIS AND EVALUATION

• That is, A gives an argument for conclusion C with reasons (also called 'premises') \( P_1 \) and \( P_2 \).

• Note, by the way, that this use of the word 'argument' has nothing directly to do with the kind of fighting argument that you might have with your roommate; rather, it’s more like the legal arguments that lawyers present to a jury.

• Also, for the sake of simplicity, I’m assuming that A gives only two reasons for believing C. In a real case, there might be only one reason (for example: Fred is a computer scientist; therefore, someone is a computer scientist), or there might be more than two reasons (for examples, see any of the arguments for analysis and evaluation in Appendix A.)

**Step 3** In order to be rational, Q should analyze or “verify” A’s arguments. Q can do this by asking three questions:

(a) Do I believe \( P_1 \)? (That is, do I agree with it?)

(b) Do I believe \( P_2 \)? (That is, do I agree with it?)

(c) Does C follow validly from \( P_1 \) and \( P_2 \)?

There are a few comments to make about Step 3:

• Strictly speaking, when you’re analyzing an argument, you need to say, for each premise, whether it is or is not true. But sometimes you don’t know; after all, truth is not a matter of logic, but of correspondence with reality (as we discussed in §2.4.1): A sentence is true if and only if it correctly describes some part of the world. (And it’s false otherwise.) Whether or not you know the truth-value of a statement (whether it’s a premise or a conclusion), you usually have some idea of whether you believe it or not. Because you can’t always or easily tell whether a sentence is true, we can relax this a bit and say that sentences can be such that either you agree with them or you don’t. So, when analyzing an argument, you can say either: “This statement is true (or false)”, or (more cautiously) “I think that this statement is true (or false)”; or “I believe (or don’t believe) this statement”, or “I agree (or don’t agree) with it”. (Of course, you should also say why you do or don’t agree!)

• Steps 3(a) and 3(b) are “recursive” (see §2.10.4); That is, for each reason \( P_i \), Q could play another instance of the game, asking A (or someone else!) whether \( P_i \) is true. A (or the other person) could then give an argument for conclusion \( P_i \) with new premises \( P_3 \) and \( P_4 \). Clearly, this process could continue. (This is what toddlers do when they continually ask their parents “Why?”). Recall our discussion of this in §2.5.1.) It is an interesting philosophical question, but fortunately beyond our present scope, to consider where, if at all, this process might stop.

• To ask whether C follows “validly” from the premises is to assume that A’s argument is a deductive one. For the sake of simplicity, all (or at least
most) of the arguments at the ends of some of the chapters are deductive. But, in real life, most arguments are not completely deductive, or not even deductive at all. So, more generally, in Step 3(c), Q should ask whether C follows rationally from the premises: If it does not follow deductively, does it follow inductively? Abductively? And so on.

- Unlike Steps 3(a) and 3(b) for considering the truth value of the premises, Step 3(c)—determining whether the relation between the premises of an argument and its conclusion is a rational one—is not similarly recursive, on pain of infinite regress.

Further Reading:
The classical source of this observation is due to Lewis Carroll (of “Alice in Wonderland” fame). (Though the books are more properly known as Alice’s Adventures in Wonderland and Through the Looking Glass.) Carroll was a logician by profession, and wrote a classic philosophy essay on this topic, involving Achilles and the Tortoise (Carroll, 1895).

- Finally, it should be pointed out that the order of doing these steps is irrelevant. Q could first analyze the validity (or rationality) of the argument and then analyze the truth value of the premises (that is, decide whether to agree with them), rather than the other way round.

Step 4 Having analyzed A’s argument, Q now has to evaluate it, by reasoning in one of the following ways;

- If I agree with P₁, 
  and if I agree with P₂, 
  and if C follows validly (or rationally) from P₁ and P₂, 
  then I logically must agree with C (that is, I ought to believe C).
  - But what if I really don’t agree with C?
    In that case, I must reconsider my having agreed with P₁, or with P₂, or with the logic of the inference from P₁&P₂ to C.

- If I agree with P₁, 
  and if I agree with P₂, 
  but the argument is invalid, is there a missing premise—an extra reason—that would validate the argument and that I would agree with?
  (See §2.10.3, below.)
  - If so, then I can accept C,
    else I should not yet reject C, 
    but I do need a new argument for C 
    (that is, a new set of reasons for believing C).

- If I disagree with P₁ or with P₂ (or both), 
  then—even if C follows validly from them—this argument is not a reason for me to believe C
  so, I need a new argument for C.
(Recall our discussion of “first-person philosophy” in §2.7.)
– There is one other option for Q in this case: Q might want to go back and reconsider the premises. Maybe Q was too hasty in rejecting them.

• What if Q cannot find a good argument for believing C? Then it might be time to consider whether C is false. In that case, Q needs to find an argument for C’s negation: Not-C (sometimes symbolized ‘¬C’).

This process of argument analysis and evaluation is summarized in the flowchart in Figure 2.6.

2.10.3 Missing Premises

One of the trickiest parts of argument analysis can be identifying missing premises. Often, this is tricky because the missing premise seems so “obvious” that you’re not even aware that it’s missing. But, equally often, it’s the missing premise that can make or break an argument.

Here’s an example from the “Textual Entailment Challenge”, a competition for computational-linguistics researchers interested in knowledge representation and information extraction. (For some real-life examples, see §§3.5, 3.13.1.2 and 5.6.2.) In a typical challenge, a system is given one or two premises and a conclusion (to use our terminology) and asked to determine whether the conclusion follows from the premise. And “follows” is taken fairly liberally to include all kinds of non-deductive inference.

Further Reading:
For more information on “textual entailment” in general, and the Challenge in particular, see Dagan et al. 2006; Bar-Haim et al. 2006; Giampiccolo et al. 2007.

Here is an example:

Premise 1 (P):
Bountiful arrived after war’s end, sailing into San Francisco Bay 21 August 1945.

Premise 2:
Bountiful was then assigned as hospital ship at Yokosuka, Japan, departing San Francisco 1 November 1945.

Conclusion (C): Bountiful reached San Francisco in August 1945.

The idea is that the two premises might be sentences from a news article, and the conclusion is something that a typical reader of the article might be expected to understand from reading it.

I hope you can agree that this conclusion does, indeed, follow from these premises. In fact, it follows from Premise 1 alone. In this case, Premise 2 is a “distractor”.

But what logical rule of inference allows us to infer C from P?

• P talks of “arrival” and “sailing into”, but C talks only of “reaching”.

• P talks of “San Francisco Bay”, but C talks only of “San Francisco”.

Further Reading:
For more information on “textual entailment” in general, and the Challenge in particular, see Dagan et al. 2006; Bar-Haim et al. 2006; Giampiccolo et al. 2007.
Figure 2.6: How to evaluate an argument from premises $P_1$ and $P_2$ to conclusion $C$. (The symbol ‘∃’ should be read: “Does there exist”.)
There are no logical rules that connect these concepts. Most people, I suspect, would think that no such rules would be needed. After all, isn’t it “obvious” that, if you arrive somewhere, then you have reached it? And isn’t it “obvious” that San Francisco Bay must be in San Francisco?

Well, maybe. But, whereas people might know these things, computers won’t, unless we tell them. In other words, computers need some lexical knowledge and some simple geographical knowledge. (If you don’t like the word ‘knowledge’ here, you can substitute ‘information’. Instead of telling the computer these additional facts, we might tell the computer how to find them; we’ll discuss these two options in §3.6.1.)

So, we need to supply some extra premises that link P with C more closely. These are the “missing premises”. The argument from P to C is called an ‘enthymeme’, because the missing premises are “in” (Greek ‘en-’) the arguer’s “mind” (Greek ‘thy-mos’).

We might flesh out the argument as follows (there are other ways to do it; this is one that comes to my mind):

\[(P) \text{ Bountiful arrived after war’s end, sailing into San Francisco Bay}
21 August 1945,]
\[(P_a) \text{ If something sails into a place, then it arrives at that place.}
(C_1) \therefore \text{ Bountiful arrived at San Francisco Bay 21 August 1945.}\]

In this first step, I’ve added a missing premise, \(P_a\), and derived an intermediate conclusion \(C_1\). Hopefully, you agree that \(C_1\) follows validly (or at least logically in some way, that is, rationally) from \(P\) and \(P_a\).

We have no way of knowing whether \(P\) is true, and must, for the sake of the argument, simply assume that it is true. (Well, we could look it up, I suppose; but we’re not investigating whether the argument is “sound” (see §2.10.4, below), only if it is “valid”: Does \(C\) follow from \(P\)?)

\(P_a\), on the other hand, doesn’t have to be accepted at all; after all, we are imposing it on the (unknown) author of the argument. So, we had better impose something that is likely to be true. \(P_a\) is offered as part of the meaning of “sail into”. I won’t defend its truth any further here, but if you think that it’s not true, then you should either reject the argument or else find a better missing premise.

We might have chosen another missing premise:

\[(P_b) \text{ If something arrives in a place named ‘X Bay’,}
\text{ then it arrives at a place named ‘X’}.
(C_2) \therefore \text{ Bountiful arrived at San Francisco 21 August 1945.}\]

\(C_2\) will follow from \(C_1\) and \(P_b\), but is \(P_b\) true? Can you think of any bays named ‘X Bay’ that are not located in a place named ‘X’? If you can, then we can’t use \(P_b\). Let’s assume the worst: Then we’ll need something more specific, such as:

\[(P_c) \text{ If something arrives in San Francisco Bay,}
\text{ then it arrives at San Francisco.}\]

\(C_2\) will follow from \(C_1\) and \(P_c\), and we can easily check the likely truth of \(P_c\) by looking at a map.
So far, so good. We’ve now got Bountiful arriving at San Francisco on 21 August 1945. But what we need is Bountiful “reaching” San Francisco in August 1945. So let’s add:

\((P_d)\) If something arrives somewhere, then it reaches that place.

Again, this is proposed as an explication of part of the meaning of ‘arrive’, and, in particular, of that part of its meaning that connects it to \(C\).

From \(P_d\) and \(C_2\), we can infer:

\((C_3)\) Bountiful reached San Francisco 21 August 1945.

Are we done? Does \(C_3 = C\)? Look at them:

\((C_3)\) Bountiful reached San Francisco 21 August 1945.

\((C)\) Bountiful reached San Francisco in August 1945.

Think like a computer! \(C_3 \neq C\). But does \(C_3\) imply \(C\)? It will, if we supply one more missing premise:

\((P_e)\) If something occurs \((on)\) DATE MONTH YEAR, then it occurs in MONTH YEAR.

And that’s true by virtue of the way (some) people talk. So, from \(P_e\) and \(C_3\), we can infer \(C\).

So, the simple argument that we started with, ignoring its irrelevant premise, becomes this rather more elaborate one:

\((P)\) Bountiful arrived after war’s end, sailing into San Francisco Bay 21 August 1945.

\((P_a)\) If something sails into a place, then it arrives at that place.

\((C_1)\) :: Bountiful arrived at San Francisco Bay 21 August 1945.

\((P_b)\) If something arrives in a place named ‘X Bay’, then it arrives at a place named ‘X’.

\((P_c)\) If something arrives in San Francisco Bay, then it arrives at San Francisco.

\((C)\) :: Bountiful arrived at San Francisco 21 August 1945.

\((P_d)\) If something arrives somewhere, then it reaches that place.

\((C_3)\) :: Bountiful reached San Francisco 21 August 1945.

\((P_e)\) If something occurs \((on)\) DATE MONTH YEAR, then it occurs in MONTH YEAR.

\((C)\) :: Bountiful reached San Francisco in August 1945.
2.10. APPENDIX: ARGUMENT ANALYSIS AND EVALUATION

2.10.4 When Is an Argument a “Good” Argument?

As we have seen, Q needs to do two things to analyze and evaluate an argument:

1. decide whether the premises are true (that is, decide whether to agree with, or believe, the premises), and
2. decide whether the inference (that is, the reasoning) from the premises to the conclusion is a valid one.

That is, there are two separate conditions for the “goodness” of an argument:

1. factual goodness: Are the premises true? (Or do you believe them?)
2. logical goodness: Is the inference valid? (Or at least rational in some way?)

Factual goodness—truth—is beyond the scope of logic, although it is definitely not beyond the scope of deciding whether to accept the conclusion of an argument. As we saw in §2.4, there are several ways of defining ‘truth’ and of determining whether a premise is true. Two of the most obvious (though not the simplest to apply!) are (a) constructing a (good!) argument for a premise whose truth value is in question and (b) making an empirical investigation to determine its truth value (for instance, performing some scientific experiments or doing some kind of scholarly research).

Logical goodness (for deductive arguments) is called ‘validity’. I will define this in a moment. But, for now, note that these two conditions must both obtain for an argument to be “really good”: A “really good” argument is said to be “sound”:

An argument is sound if and only if it is both valid and “factually good”, that is, if and only if it is both valid and all of its premises really are true.

Just to drive this point home: If the premises of an argument are all true (or if you believe all of them)—and even if the conclusion is also true—that by itself does not make the argument sound (“really good”). For one thing, your belief in the truth of the premises might be mistaken. But, more importantly, the argument might not be valid.

And, if an argument is valid—even if you have doubts about some of the premises—that by itself does not make the argument sound (“really good”). All of its premises also need to be true; that is, it needs to be factually good.

So, what does it mean for a (deductive) argument to be “valid”?

An argument is valid if and only if it is necessarily “truth-preserving”.

Here’s another way to put it:

An argument is valid
if and only if
whenever all of its premises are true, then its conclusion must also be true.

And here’s still another way to say the same thing:

An argument is valid
if and only if
it is impossible that:
all of its premises are simultaneously true while its conclusion is false.
Note that this has nothing to do with whether any of the premises actually are true or false; it’s a “what if” kind of situation. Validity only requires that, if the premises were to be true, then the conclusion would preserve that truth—it would “inherit” that truth from the premises—and so it would also (have to) be true.

So you can have an argument with false premises and a false conclusion that is invalid, and you can have one with false premises and a false conclusion that is valid. Here’s a valid one:

All cats are fish.
All fish can fly.
\[ \therefore \text{All cats can fly.} \]

Here, everything’s false, but the argument is valid, because it has the form:

\[ \begin{align*}
\text{All } P \text{s are } Q \text{s.} \\
\text{All } Q \text{s are } R \text{s.} \\
\therefore \text{All } P \text{s are } R \text{s.}
\end{align*} \]

and there’s no way for a \( P \) to be a \( Q \), and a \( Q \) to be an \( R \), without having the \( P \) be an \( R \). That is, it’s impossible that the premises are true while the conclusion is false.

Here’s an invalid one, also with false premises and conclusion:

All cats are fish.
All cats can fly.
\[ \therefore \text{All fish can fly.} \]

Again, everything’s false. However, the argument is invalid, because it has the form:

\[ \begin{align*}
\text{All } P \text{s are } Q \text{s.} \\
\text{All } P \text{s are } R \text{s.} \\
\therefore \text{All } Q \text{s are } R \text{s.}
\end{align*} \]

and arguments of this form can have true premises with false conclusions. Here is an example:

All cats are mammals.
All cats purr.
\[ \therefore \text{All mammals purr.} \]

So, it’s possible for an argument of this form to have true premises and a false conclusion; hence, it’s not valid.

To repeat: Validity has nothing to do with the actual truth or falsity of the premises or conclusion. It only has to do with the relationship of the conclusion to the premises. Recall that an argument is sound iff it is valid and all of its premises are true. Therefore, an argument is unsound iff either it is invalid or at least one premise is false (or both). An unsound argument can be valid!

One more point: An argument with inconsistent premises (that is, premises that contradict each other) is always valid(!), because it’s impossible for it to have all true premises with a false conclusion, and that’s because it’s impossible for it to have all
true premises, period. Of course, such an argument cannot be sound. (The argument that Bertrand Russell is the Pope that we saw in §2.6.1.1 is an example of this.)

All of this is fine as far as it goes, but it really isn’t very helpful in deciding whether an argument really is valid. How can you tell if an argument is truth-preserving? There is a simple, recursive definition, but, to state it, we’ll need to be a bit more precise in how we define an argument.

**Definition 1:**
An argument from propositions $P_1, \ldots, P_n$ to conclusion $C$ is defined a sequence of propositions $(P_1, \ldots, P_n, C)$, where $C$ is alleged to follow logically from the $P_i$.

**Definition 2:**
An argument $(P_1, \ldots, P_n, C)$ is valid if and only if each proposition $P_i$ and conclusion $C$ is either:

(a) a tautology

(b) a premise

(c) or follows validly from previous propositions in the sequence by one or more truth-preserving “rules of inference”.

This needs some commenting! (a) First, a tautology is a proposition that must always be true. How can that be? Most tautologies are (uninformative) “logical truths”, such as ‘Either $P$ or not-$P$’, or ‘If $P$, then $P’. Note that, if $P$ is true (or, if you believe $P$), then ‘Either $P$ or not-$P$’ has to be true (or, you are logically obligated to also believe ‘Either $P$ or not-$P$’), and, if $P$ is false, then not-$P$ is true, and so ‘Either $P$ or not-$P$’ still has to be true (or, you are logically obligated to also believe ‘Either $P$ or not-$P$’). So, in either case, the disjunction has to be true (or, you are logically obligated to believe it). Similar considerations hold for ‘If $P$, then $P’. Sometimes, statements of mathematics are also considered to be tautologies (whether they are “informative” or not is an interesting philosophical puzzle; see Wittgenstein 1921).

(b) Second, a premise is one of the initial reasons given for $C$, or one of the missing premises added later. Premises, of course, need not be true, but, when evaluating an argument for validity, we must assume that they are true “for the sake of the argument”. Of course, if a premise is false, then the argument is unsound.

Third, clause (c) of Definition 2 might look circular, but it isn’t; rather, it’s recursive. A “recursive” definition begins with “base” cases that give explicit examples of the concept being defined and then “recursive” cases that define new occurrences of the concept in terms of previously defined ones. (We’ll say a lot more about recursion in Chapter 7.)

In fact, this entire definition is recursive. The base cases of the recursion are the first two clauses: Tautologies must be true, and premises are assumed to be true. The recursive case consists of “rules of inference”, which are argument forms that are clearly valid (truth-preserving) when analyzed by means of truth tables.

So, what are these “primitive” valid argument forms known as “rules of inference”? The most famous is called ‘Modus Ponens’ (recall §2.6.1.1):

---

6This symbol means “is by definition”.
From $P$ and ‘If $P$, then $C$’, you may validly infer $C$.

Why may you validly infer $C$? Consider the truth table for ‘If $P$, then $C$’:

<table>
<thead>
<tr>
<th>$P$</th>
<th>$C$</th>
<th>If $P$, then $C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>true</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>true</td>
<td>false</td>
<td>false</td>
</tr>
<tr>
<td>false</td>
<td>true</td>
<td>true</td>
</tr>
<tr>
<td>false</td>
<td>false</td>
<td>true</td>
</tr>
</tbody>
</table>

It says that that conditional proposition is false in only one circumstance: when its “antecedent” ($P$) is true and its “consequent” ($C$) is false. In all other circumstances, the conditional proposition is true. So, if the antecedent of a conditional is true, and the conditional itself is true, then its consequent must also be true. (Look at the first line of the truth table.) Modus Ponens preserves truth.

Another important rule of inference is called ‘Universal Elimination’ (or ‘Universal Instantiation’):

From ‘For all $x$, $F(x)$’ (that is, for all $x$, $x$ has property $F$), you may validly infer $F(a)$, for any individual $a$ in the “domain of discourse” (that is, in the set of things that you are talking about).

A truth-table analysis won’t help here, because this is a rule of inference from “first-order predicate logic”, not from “propositional logic”. The formal definition of truth for first-order predicate logic is beyond our scope, but it should be pretty obvious that, if it is true that everything in the domain of discourse has some property $F$, then it must also be true that any particular thing in the domain (say, $a$) has that property. (For more rules of inference and for the formal definition of truth in first-order predicate logic, see any good introductory logic text or the Further Reading on the correspondence theory of truth, in §2.4.1, above.)

There are, however, a few terminological points to keep in mind:

- **Sentences** can only be true or false (or you can agree or disagree with them).

- **Arguments** (which are sequences of sentences) can be valid or invalid, and they can be sound or unsound.

- **Conclusions** of arguments (which are sentences) can follow validly or not follow validly from the premises of an argument.

Therefore:

- **Sentences** (including premises and conclusions) cannot be valid, invalid, sound, or unsound (because they are not arguments).

- **Arguments cannot** be true or false (because they are not sentences).
2.10.5 Examples of Good and Bad Arguments

There is only one way to have a sound argument: It must be valid and have only true premises. But there are lots of ways to have invalid arguments! (For an example of one, see Figure 2.5.) More importantly, it is possible to have an invalid argument whose conclusion is true! Here’s an example:

All birds fly. (true)
Tweety the canary flies. (true)
Therefore, Tweety is a bird. (true)

This is invalid, despite the fact that both of the premises as well as the conclusion are all true (but see §19.4.3): It is invalid, because an argument with the same form can have true premises and a false conclusion. Here is the form of that argument:

\[ \forall x (B(x) \rightarrow F(x)) \]
\[ C(a) \land F(a) \]
\[ \therefore B(a) \]

In English, this argument’s form is:

For all x, if x has property B, then x has property F.
a has property C, and a has property F.
\therefore a has property B.

That is,

For all x, if x is a bird, then x flies.
Tweety is a canary, and Tweety flies.
Therefore, Tweety is a bird.

Here’s a counterexample, that is, an argument with this form that has true premises but a false conclusion:

All birds fly. (true)
Bob the bat flies. (true)
Therefore, Bob is a bird. (false)

Just having a true conclusion doesn’t make an argument valid. And such an argument doesn’t prove its conclusion (even though the conclusion is true).

Here is a collection of valid (V), invalid (I), sound (S), and unsound (U) arguments with different combinations of true (T) and false (F) premises and conclusions. Make sure that you understand why each argument below is valid, invalid, sound, or unsound.
CHAPTER 2. WHAT IS PHILOSOPHY?

A (1) All pianists are musicians. T
(2) Lang Lang is a pianist. T V S
(3) ∴ Lang Lang is a musician. T

B (1) All pianists are musicians. T
(2) Lang Lang is a pianist. T I U
(3) ∴ Lang Lang is a pianist. T

C (1) All musicians are pianists. F
(2) The violinist Itzhak Perlman is a musician. T V U
(3) ∴ Itzhak Perlman is a pianist. F

D (1) All musicians are pianists. F
(2) Itzhak Perlman is a violinist. T I U
(3) ∴ Itzhak Perlman is a pianist. F

E (1) All cats are dogs. F
(2) All dogs are mammals. T V U
(3) ∴ All cats are mammals. T

F (1) All cats are dogs. F
(2) All cats are mammals. T I U
(3) ∴ All dogs are mammals. T

G (1) All cats are dogs. F
(2) Snoopy is a cat. F V U
(3) ∴ Snoopy is a dog. T

H (1) All cats are birds. F
(2) Snoopy is a cat. F I U
(3) ∴ Snoopy is a dog. T

I (1) All cats are birds. F
(2) All birds are dogs. F V U
(3) ∴ All cats are dogs. F

J (1) All cats are birds. F
(2) All dogs are birds. F I U
(3) ∴ All cats are dogs. F

K (1) All cats are mammals. T
(2) All dogs are mammals. T I U
(3) ∴ All cats are dogs. F
2.10. APPENDIX: ARGUMENT ANALYSIS AND EVALUATION

2.10.6 Summary

So, to analyze an argument, you must identify its premises and conclusion, and supply any missing premises to help make it valid. To evaluate the argument, you should then determine whether it is valid (that is, truth preserving), and decide whether you agree with its premises.

If you agree with the premises of a valid argument, then you are logically obligated to believe its conclusion. If you don’t believe its conclusion, even after your analysis and evaluation, then you need to revisit both your evaluation of its validity (maybe you erred in determining its validity) as well as your agreement with its premises: If you really disagree with the conclusion of a valid argument, then you must (logically) disagree with at least one of its premises.

You should be sure to use the technical terms correctly: You need to distinguish between premises—which can be true or false (but cannot be “valid”, “invalid”, “sound”, or “unsound”)—and arguments—which can be valid (if its conclusion must be true whenever its premises are true), invalid (that is, not valid; its conclusion could be false even if its premises are true), sound (if it’s valid and all of its premises are true) or unsound (that is, not sound: either invalid or else valid-with-at-least-one-false-premise) (but cannot be “true” or “false”).

And you should avoid using such non-technical (hence ambiguous) terms as ‘correct’, ‘incorrect’, ‘right’, or ‘wrong’. You also have to be careful about calling a conclusion “valid”, because that’s ambiguous between meaning that you think it’s true (and are misusing the word ‘valid’) and meaning that you think that it follows validly from the premises.

You should try your hand at analyzing and evaluating the much more complex arguments in Appendix A!
Digression: Can any proposition (or its negation) be proved?
That is, given a proposition $P$, we know that either $P$ is true or else $P$ is false (that is, that $\neg P$ is true). So, whichever one is true should be provable. Is it? Not necessarily!

First, there are propositions whose truth value we don’t know yet. For example, no one knows (yet) if Goldbach’s Conjecture is true. Goldbach’s Conjecture says that all positive even integers are the sum of 2 primes; for example, $28 = 5 + 23$. For another example, no one knows (yet) if the Twin Prime Conjecture is true. The Twin Prime Conjecture says that there are an infinite number of “twin” primes, that is, primes $m,n$ such that $n = m + 2$; for example, 2 and 3, 3 and 5, 5 and 7, 9 and 11, 11 and 13, etc.

Second—and much more astounding than our mere inability so far to prove or disprove any of these conjectures—there are propositions whose truth value is known to be true, but which we can prove that we cannot prove! This is the essence of Gödel’s Incompleteness Theorem. Stated informally, it asks us to consider this proposition, which is a slight variation on the Liar Paradox (that is the proposition “This proposition is false”: If it’s false, then it’s true; if it’s true then it’s false):

$$(G) \text{ This proposition (G) is true but unprovable.}$$

We can assume that $(G)$ is either true or else false. So, suppose that it is false. Then it was wrong when it said that it was unprovable; so, it is provable. But any provable proposition has to be true (because valid proofs are truth-preserving). That’s a contradiction, so our assumption that it is false was wrong: It isn’t false. But, if it isn’t false, then it must be true. But if it’s true, then—as it says—it’s unprovable. End of story; no paradox!

So, $(G)$ (more precisely, its formal counterpart) is an example of a true proposition that cannot be proved. Moreover, the logician Kurt Gödel showed that some of them are propositions that are true in the mathematical system consisting of first-order predicate logic plus Peano’s axioms for the natural numbers (which we’ll discuss in §7.7.2.1); that is, they are true propositions of arithmetic! For more information on Gödel and his proof, see Nagel et al. 2001; Hofstadter 1979; Franzén 2005; Goldstein 2006.

We’ll return to this question, also known as the “Decision Problem”, beginning in §6.6.
Bibliography


BIBLIOGRAPHY


O’Toole, G. (2016, 5 January). In a woman the flesh must be like marble; in a statue the marble must be like flesh. *Quote Investigator*, urlhttp://quoteinvestigator.com/2017/01/05/marble/.


Schwitzgebel, E. (9 January 2012). For all x, there’s philosophy of x. http://schwitzsplinters.blogspot.com/2012/01/for-all-x-there-is-philosophy-of-x.html.


