Logic: A Colorful Approach

Shahin Kaveh

Contents

Preface ………………………………………………………………….... v

Chapter 1. Introduction: Three Forms of Reasoning …………………… 1

Chapter 2. Abductive Reasoning …………………………………………… 19

Chapter 3. Inductive Reasoning ………………………………….………..… 31

Chapter 4. Propositional Logic: Symbolism ……………….…………….… 45

Chapter 5. Propositional Logic: Truth Tables ...……………….…………..… 69

Chapter 6. Propositional Logic: Proofs ……......……………….…………… 85

Chapter 7. Quantificational Logic: Symbolism ..……..…….….………..… 117

Chapter 8. Quantificational Logic: Proofs …….……...……….………..… 133

Chapter 9. Fallacies ……………………….……..……..…….….………..… 153

Preface

I am one of those people whose brain involuntarily associates the letters of the alphabet (and therefore words) with certain colors. So, for example, the word "deductive" is kind of green to me, while the word "inductive" is red and "abductive" is blue. That means a book that covers only deductive logic is an all green book. But a book, like this one, that covers deductive, inductive, and abductive logic, as well as fallacies is a colorful book! We will learn in the introduction what these words mean, but for now you can think of them as different modes of reasoning. Most introductory logic classes focus on the mode of reasoning that we call deduction, and we are going to do the same. But we will have a quick look at induction and abduction as well. Fallacies are forms of bad argumentation that a rational person must try to avoid. We will talk about those too. But there are even more layers of color in this book, as each topic covers a variety of sub-topics. In deductive logic, for instance, we will learn about translations, truth tables, and proofs, each of which has a different feel (and color) to it.

This book has two types of practice questions. One is traditional questions on paper that you will find at the end of each chapter. But you will also be asked to work with the software LogiCola, written by Harry J. Gensler, philosopher from John Carroll University. The software is easy to install and work with and can be found at either of the following addresses:

http://www.routledge.com/textbooks/gensler_logic

http://www.jcu.edu/philosophy/gensler/logicola.htm

This software provides you with a new way of doing exercise, kind of similar to learning from flash cards: you have to keep going until you get it right. I find this a nice supplement to the traditional way.

Although this book covers everything that you will need for this course, supplemental readings are suggested at the end of each section. You are advised to read those as well, especially if you are having difficulty understanding the material. (Note to my students: These supplemental readings are available in my office and I can make copies of them for you.)

Chapter 1

Introduction: Three Forms of Reasoning

§1.1 What is Logic?

The study of logic can be helpful for all people, including math and science majors. Sciences like physics and chemistry make extensive use of logic in their calculations and theorems, while biology, psychology, sociology, and medicine make use of logic in developing and critically assessing theories. All of these disciplines also use logic when they interpret the data collected in the laboratory or in the surveys or medical controlled trials.

But what is logic? Logic is the study of human reasoning. This is a very rough definition and might not distinguish logic from many other disciplines. For example, you might ask yourself "Isn't psychology supposed to study human reasoning?". To be sure, if we are interested in the question of how people actually reason, what does and does not convince them, and what goes on in their mind when they infer something from something else, the discipline to turn to for answers is psychology. However, what kinds of reasoning are appropriate and correct and what kinds are inappropriate and flawed seems to be independent of human psychology. People are sometimes convinced by flawed arguments. The psychologist can tell us how many, in what circumstances, and maybe why people are convinced by bad arguments. But the logician can tell us why an argument is good or bad. What is amazing about logic is that, when you bring those very people who are convinced by a flawed argument and tell them why it is flawed, the vast majority of them see through their mistake and accept the logician's point. Thus, it seems that logic is something shared by almost all human beings, while it is independent of psychology. It is sometimes assumed that all people already know logic, even if they are not aware of it. But studying logic can sharpen this inchoate knowledge and make it more conscious and explicit and provide a common language to communicate our critiques.

To begin, let me introduce some terminology. Logicians distinguish what they call "statements", also called "propositions", from other kinds of sentences. A *statement* or *proposition* is a collection of words, such as "Clinton is not a Republican", that is capable of being *true* or *false*. A phrase like "Vote for Clinton!" is not a proposition, because it cannot be true or false. It is simply a command, which might be a good or bad one, but in any case it cannot be true or false. For the same reason, a question like "Did you vote for Clinton?" is not a statement/proposition. Logicians also distinguish between propositions and arguments. An *argument* consists of a series of claims which are supposed to *establish* that a certain statement is true. When your friend does not believe something that you believe, normally you will not just keep *asserting* that your claim is true. Rather, you use an argument to *convince* them. For example, the following is an argument:

1) All Republicans are against government control of the market.

2) Clinton is not a Republican.

---

3) Therefore, Clinton is not against government control of the market.

Lines 1 and 2 are called the *premises* of the argument, which is the logician's word for the *assumptions*, or the evidence provided. The sentence that is being argued for (such as 3) is called the *conclusion*. The horizontal line separates the conclusion from the premises. The premises of an argument are claimed to *support* the conclusion. We will see below that this support can come in a variety of ways. The premises and the conclusion are of course propositions and might be true or false. But the argument itself, as a process of reasoning, *cannot be true or false*. It's just not the type of thing that can be true or false. The properties that can be rightly attributed to an argument are *good* and *bad* (the same way a command cannot be true or false but can be good or bad). A good argument is one in which the premises provide strong support for the conclusion. A bad argument is one in which that is not the case. In a moment, I will introduce more accurate terminology for good and bad arguments.

The argument given above is flawed (you'll see why below). You might notice that all three sentences in this argument are, most probably, true; so the flaw is not in saying something false or dubious, it is elsewhere. If you find it surprising that this argument is bad, it is because this is one of the examples of bad arguments that can sometimes convince people due to psychological effects. Unlike psychology, logic teaches us to see the flaw in this argument and hopefully makes us less prone to making such mistakes.

It is often said that logic as a discipline was invented by Aristotle (384-322 BC). He identified several forms of good and bad reasoning and formulated them in the form of what came to be known as "syllogisms". An example of one type of syllogism is 1-3 above. Aristotle was especially interested in syllogisms partly because the premises of a good syllogism not only support but *guarantee* the conclusion. This is the strongest and most well-known form of support that the premises can provide for the conclusion. We will see below that an argument can be a good argument without having this feature: the premises can give us *some* evidence to believe in the conclusion, while not guaranteeing it 100%. Aristotle's most significant discovery was the following: Whether the premises of an argument *guarantee* its conclusion has nothing to do with the *content* of what is being said, but has everything to do with the *form* or *structure* of the argument. All that matters is how the different elements in the argument are arranged with respect to each other. What each of these elements says is not important. Let's see what that means through the example made above. You can see that the phrases “Republican”, “Clinton”, and “against government control of the market” occur in various places in 1-3 above. We can separate the form and the content by symbolizing the argument this way:

1') All R is A.

2') c is not R.

3') Therefore, c is not A.

We have replaced the English phrases with letters (you might notice that some of the letters are uppercase and some are lowercase; don’t worry about that just yet). We said before and will see later on that this argument is flawed. For now just accept from me that it is a bad argument. Now, Aristotle’s great discovery was this: as far as reasoning is concerned, *it does not matter* what R, A, and c stand for. They can refer to anything or anyone; the argument will still be a flawed one in the sense that the premises don’t guarantee the conclusion. In other words, *any argument that has the above form* is a flawed argument. So, for example, if R stands for “raven”, A stands for “able to fly” and c stands for Peter Pan, we know the following argument is also a bad argument:

1") All ravens are able to fly.

2") Peter Pan is not a raven.

3") Therefore, Peter Pan is not able to fly.

The same is true of good arguments. For example, here is a famous example used by Aristotle himself:

4) All humans are mortal.

5) Socrates is a human.

---

6) Therefore, Socrates is mortal.

It is hard to deny that this is good reasoning (I hope your intuitions are clear on this one). Now, the point is that any other argument that has the above form (what is its form?) is also a good argument, *regardless* of its content. Consider this one for example:

4') All horses are yellow.

5') The Earth is a horse.

6') Therefore, the Earth is yellow.

This argument is identical in form to 4-6 and therefore has to be a good argument. Now you might wonder “Is it really?”. There is certainly something really wrong about these sentences. On the other hand, there is something legitimate about the flow of the information from 4' and 5' to 6'. To be sure, some of these sentences are false or maybe even nonsensical, but the *structure* of the argument is a good one. What this means is that *if* someone were to accept that 4' and 5' are true, he/she would *have to* accept that 6' is true, on pain of rationality. This is the sign of a good argument. To avoid confusion, let us introduce two other pieces of very important terminology: We call an argument *valid* if the truth of the premises guarantees the conclusion to be true. In other words, it is not possible that the premises are true and the conclusion is false. We call an argument *sound* if it is valid *and* its premises are actually true. So, for example, 4-6 is valid and sound, while 4'-6' is valid and unsound. Aristotle’s major discovery was that the validity of an argument is purely determined by its form and not by its content. We shall talk a lot more about these concepts later on.

Premises:

The assumptions; the evidence provided

Conclusion:

The sentence being argued for; the final assertion

Valid:

No possibility that the premises are true and the conclusion is false

Sound:

Valid with true premises

Why is Aristotle’s discovery important? It is important because there are amazingly few forms of valid arguments, while there are infinitely many *invalid* ones. Thus, once we realize that validity is all about the form, the study of good reasoning can be reduced to the study of those few valid forms. Everything else will be bad reasoning.

You can already see why the argument 1-3 at the beginning of this chapter was flawed. It is possible that someone is not a Republican but is opposed to government control of the market for other reasons. Therefore, it is *possible* that the premises, i.e. propositions 1 and 2, are true, but the conclusion, i.e. proposition 3, is false (whether it is *actually* true or false has nothing to do with the validity of the argument). So 1-3 is an invalid and unsound argument.

For the longest time, logic was little more than what Aristotle had written. In the 19th century, however, logicians realized that there are other forms of valid reasoning that Aristotle did not mention. The German mathematician and philosopher, Gottlob Frege, created a new method of formalizing these. This led to an explosion of systems of logic, which is why Aristotelian syllogisms are today a small part of the discipline of logic. But the basic idea remained the same. We are looking for valid forms. Here is one valid argument that Aristotle's logic did not cover:

7) If John attends the meeting, Mary attends the meeting too.

8) John will attend the meeting.

---

9) Therefore, Mary will attend the meeting.

The form of the argument is:

7') If P, then Q

8') P

9') Therefore, Q

Since there is no way that the premises can be true without the conclusion also being true, this is a valid argument (and a very famous and intuitive one). If, in addition, the premises are in fact true, the argument is sound. In contrast with 7-9, consider this other argument:

10) If you work hard, you become rich.

11) You are rich.

---

12) Therefore, you have worked hard.

It is possible that one becomes rich without working hard, for example by inheriting money from his/her parents. Therefore, it is possible that the premises are true and yet the conclusion is false. So the argument is invalid and therefore automatically unsound. The form of this (invalid) argument is shown below. Note carefully how it is different from 7'-9'.

10') If P, then Q

11') Q

12') Therefore, P

When the premises are claimed to guarantee the conclusion, the argument is called a *deduction*. The discipline that deals with this kind of arguments is *deductive logic*. Of course, by learning the valid forms of reasoning, we also learn about the *invalid* forms. So, deductive logic deals with all arguments in which the premises are claimed to provide 100% support for the conclusion and tells us which ones are valid and which ones are not.

It has always been known that valid arguments are not the only type of argument that can be called good. There are good arguments used in everyday life, in science, and in court of law that are not valid in the sense defined above; that is to say, accepting the truth of the premises does not *force* you to accept the conclusion. But these arguments can still be good in the sense that accepting the premises gives you *some* reason to accept the conclusion. These reasons might sometimes be pretty strong even if they never live up to the airtight standards of a deduction. Two such forms of reasoning that we will discuss shortly are called *induction* and *abduction*. Some people reserve the label “logic” for deductive logic. But we will use it more broadly in this course, to cover inductive and abductive logic as well. Also note that many logicians use “induction” as a label for any kind of reasoning that is not deductive (so all arguments will be either deductive or inductive in this terminology). In this course, we will preserve “induction” for a specific class of non-deductive arguments that I shall explain shortly.

There has been a great deal of interesting and ongoing discussion in philosophy about whether an argument that is not deductively valid (such as an induction or abduction) is any good at all. In this course, we will not go into any of these debates. We shall simply assume that inductions and abductions can be good reasoning provided they are done well, and merely concern ourselves with the question of *how* they can be done well. But first let me clarify what inductions and abductions are by comparing them to deductions.

§1.2 Deductions, Inductions, Abductions, and Observations

§1.2.1 Deductions

Let me recap what I said of deductions above. Deductive arguments are those arguments of which we can say that if they are good it is because they are valid and if they are bad it is because they are invalid, where *validity* is defined as follows:

> *An argument is valid if and only if there is no possibility that the premises are true and the conclusion is false.*

We also said that validity is all about the form:

> *The validity or invalidity of an argument is exclusively determined by its form.*

Since in deductive logic good equals valid, we can say that deductive arguments have the two following properties:

> D1) A good deductive argument is one in which there is no possibility that the premises are true and the conclusion is false.
>
> D2) Whether a given deductive argument is good is exclusively determined by its form.

It is very important to remember that D1 does *not* say that the conclusion of the argument must be or is true. Nor does it say that if everything you say is true, then your argument is valid. A valid argument can have false or true premises as well as a false or true conclusion (remember 4'-6'). The only thing it *cannot* have is all true premises and a false conclusion. D2 helps you remember this because when you abstract away from the content and replace the sentences with symbols, you cannot tell whether the premises are true or not, but you can still tell that *if* the premises are true, the conclusion has to be true.

D1 and D2 are really neat features for an argument to have. Every time you hear an argument, you can ask the speaker to put the argument in a deductively valid form and then you would not have to worry about anything other than whether each premise is true. So the entire discussion can come down to a couple pieces of information regarding the truth of the premises. If, on the other hand, the person cannot put the argument in a valid format, the reason is that one or more of his/her assumptions are not explicitly mentioned and the argument cannot be valid without them (these are called "hidden assumptions"). People who are not familiar with this method often get into long and confused discussions where it is not clear what is being disagreed on. Is it the truth of the assumptions involved or whether the assumptions properly support the conclusion? Putting the argument in the form of a deduction nicely separates these two issues from each other. We can then be 100% certain that the conclusion follows from the premises.

This absolute certainty has fascinated people for centuries, which is why many philosophers and scientists have tried to create systems that imitate mathematics (because mathematics is all deductions). But this certainty and elegance comes at the following price.

D3) The conclusion of a deduction contains no more information than the premises.

You might have noticed this already. The conclusions of the arguments we discussed above seem to just restate the information in the premises in a different way (think of 4-6 for example). This is true of all deductions. Deductions serve to rearrange, reformulate, or clarify our information, but they don't add to it. However, note that this is not always obvious. Sometimes you might feel like you have learned something new when you hear the argument. Sometimes you might even be surprised by the conclusion. All mathematical proofs are deductions but the results are not always simple or obvious. Some of them are very complicated or surprising or both. But we still say that, logically, the conclusion contains no more information than the premises. Consider the following argument, for example:

13) Everyone loves lovers.

14) John loves Mary.

---

15) Therefore, everyone loves everyone.[1]

The conclusion of this argument does not seem at first to follow from the premises. So, you might think that the argument is invalid. But in fact we can use logical calculations to show that it is valid. Thus, the conclusion is surprising and sounds like some new information that was not contained in the premises. But consider this reasoning: since John loves Mary, John is a lover. Therefore, we can conclude from 13 that everyone loves John. But then everyone would be a

[1] This example is taken from *The Logic Book*, Bergmann et. Al., McGraw-Hill, 2009, p. 10

John-lover and therefore a lover, and so again from 13 we know that everyone loves everyone else. By thinking about this step-by-step reasoning, we can come to realize that the conclusion *was* in fact contained in the premises, if not in an obvious way. So this fact about information containment does not mean that deductions can't be helpful or fun.

So, to summarize, deductions have three important properties: they are good if and only if they are valid, whether they are good or bad only depends on their form, and finally, their conclusions contain no more information than their premises.

§1.2.2 Inductions

Every time you say that you "learned from experience" you are using an *induction*. So, for example, if you take several chemistry classes and fail all of them, you might conclude that you will always fail chemistry. If we put your (implicit) reasoning in the premise-conclusion form, it would be something like this:

I failed chemistry $time_1$.

I failed chemistry $time_2$.

I failed chemistry $time_3$.

.

.

.

Therefore, I probably fail chemistry every time.

More generally, an induction is an argument in which the premises describe what happened in certain times and places in the past and the conclusion generalizes that the same thing will happen in every time and every place. Many of the things that we firmly believe in, such as that the sun will rise tomorrow or that water boils at 100 degrees Celsius, are the results of inductions. Stereotypes such as "all white people are racist" are also often inductions. You can see that there are good inductions and bad inductions.

Medical statements are also usually inductions. When the doctor tells you that smoking increases the chance of cancer she is making an inductive argument. However, the way induction is done in science is much more sophisticated than in everyday life. That is part of why their inductions are often better than the casual inductions we use everyday. We will work with several examples in Chapter 3.

Inductions do not have the properties that we labeled D1-D3 above. That they do not have D1 and D2 is kind of bad news, while that they do not have D3 is good. Let us see what that means. First of all, no matter how good or bad an induction is, it is never going to be a *valid* argument in the sense defined above. In other words, no matter how many times you have seen water boil at 100 degrees Celsius, you cannot be 100% sure that it will boil at the same temperature next time. But perhaps you can be fairly sure, if you have observed this phenomenon many times and under many different circumstances. (Some philosophers such as David Hume doubt this, but remember that in this course we are not planning to go into these philosophical issues.) This automatically means that inductions don't have D2 either. It *does* matter what the content of the argument is, because if we symbolize the above argument we will end up with something like this:

A

B

C

.

.

.

Therefore, Z

Obviously, not every argument with this (rather general) form is going to be a good argument. It depends on what A, B, etc. say. You might think that the $time_1$, $time_2$ stuff matters and should not be omitted when we extract the form. But the following example shows that the problem is deeper than that:

I woke up $time_1$ and it was the 20th century.

I woke up $time_2$ and it was the 20th century.

.

.

.

Therefore, it will always be the 20th century.

This won't be a good argument no matter how many times and under what variety of circumstances you have made your observations. Obviously, one day it is guaranteed to be the 21st century.

So, to summarize, inductions are generalizations based on past experience, they can be *strong* or *weak*, but they can never be valid. But there is some good news too. As you can see in the above examples, the conclusion of an induction does go beyond the information contained in the premises. The sentence that "smoking always increases the chance of cancer" says a lot more than all the premises combined: "smoking increased the chance of cancer in $experiment_1$", "smoking increased the chance of cancer in $experiment_2$", etc. The experiments mentioned in the premises probably involved no more than a few hundred or thousand people, but the conclusion is about all people in the world in all times and places. So, you can see that logically, there is a trade-off between certainty and gain in information, as it were. Deductions give us absolute certainty in reasoning but no new information, while inductions increase our knowledge at the price of some uncertainty. We shall see that the same is true of abductions.

§1.2.3 Abductions

The third and last form of arguments that we will discuss is abduction. This subject (but not the argument itself) is indeed the youngest of all, for philosophers did not seriously discuss abductive reasoning until very recently. Suppose you are walking on the beach and see footprints. Most of us immediately think that someone has been walking on the sand. How did we come to this belief? We did not *see* the person walking there, so our belief is not the result of an immediate observation. That someone has been walking there might not be a generalization from past experience either. It might be your first time walking on the beach and your first time seeing footprints. So your reasoning is probably not an induction either. You might want to formulate it as a deduction as follows:

16) If there are footprints on the sand, someone has been walking there.

17) There are foot prints on the sand.

---

18) Therefore, someone has been walking there.

But in this situation the interesting question is how do you know 16 is true? The answer to this question can no longer be deduction or induction or observation (why not?). So it has to be something else. It seems that someone having walked on the sand is a very good *explanation* of what you are observing (i.e. the footprints). So perhaps there is a form of reasoning from observations to explanations? That is exactly what abduction is about. You are using abductive reasoning every time you think of a good explanation for what you see. But we immediately face a problem: there are infinitely many possible explanations for any set of observations. In this case, for example, your observations are:

O1) There are footprints on the sand.

O2) The footprints have the shape of human shoes.

O3) The footprints are regular.

O4) Everything else seems to be normal.

This list of observations can be extended indefinitely, but let us stick to these four. Now, one can think of many possible explanations here. For instance:

E1) Someone has been walking here.

E2) A monkey with human shoes has been walking here.

E3) Aliens have created these marks with some high technology to deceive me.

E4) Through a rare but possible process, the wind has accidentally created these marks on the sand.

E5) Some invisible ghost has made the marks on the sand. The ghost is extremely careful not to be seen or sensed and will therefore never create any hint while you are watching, but it does things when no one is paying attention.

These are all possible explanations for your observations O1-O4. But I hope you agree that the *best* explanation based on the given information is E1. There is something fishy about each of E2-E4. For example, E2 and E4 are unlikely scenarios. E4 is of course much *more* unlikely. Also, E2 and E3 seem to be unnecessarily complicated stories when we can explain the whole thing with much simpler assumptions such as E1. Moreover, E3 is also problematic because there is a well-established, scientifically supported belief in the belief system of most of us that there has never been good evidence for the existence of aliens near the Earth. E5 seems to have another problem. There is no way to test the theory or investigate into its claims any further. We will talk about all these criteria for finding the best explanation in Chapter 2, but I hope you already have a rough picture of how abduction works. Our minds gather data by *observing* things. We then think of some possible explanations and pick the one that is best, i.e. the simplest, most testable, most likely scenario that is also compatible with our well-established beliefs. For this reason, another name for abduction is *inference to the best explanation*.

Abductions are more common than you might think. We use abductions all the time. For instance, when you call or visit your friend and they don't answer the phone or the door, your thought that "they are not home" is an inference to the best explanation of your observations. Notice that you don't *see* that they are not home, so it is not a direct observation. It is very common to talk about abductions as if they are observations. For instance, people say things like "so and so hates me" as if they can see the hatred. But really the hatred is simply an explanation for the other things that are actually observed, such as how they act or talk or treat them. Even your own emotions are often unavailable for introspection. But since your brain constantly makes "guesses" about how you are feeling, you might think that you can observe your own emotions. Here are some more examples of everyday abductions. You can't find your wallet and think that you must have dropped it somewhere on the way. You call your boyfriend several times and he doesn't pick up the phone. You email him and he doesn't reply. You remember that he used to pick up the phone or reply back very quickly in the past. You also remember that recently you said something that he didn't like. You think he must be mad at you. Abduction is also very common in science. The most obvious examples might be from police investigations (or detective stories) and archaeology. But all sciences use abduction. The theory of evolution, for instance, is the best explanation (as of today) of all the observations we have of various species, fossils, and several other biological and geological processes. Long before we could actually see the Earth from outside, we inferred that it is round and turns around the Sun, because that was the best explanation of a wide array of observations on Earth. The claim that a particular collection of molecules called chromosomes determines a significant portion of every animal's physical and mental features is a hypothesis widely believed to be the best explanation of various sorts of observations in biology. It is instructive to think about each of these scenarios more carefully. Ask yourself what the relevant observations are. Try to carefully distinguish what is a genuine observation from what is merely inferred from them. And finally, find a few alternative explanations and ask yourself why they are worse explanations. We will study several cases like these later on.

Abductions also lack all of the major properties of deductions that allowed us to apply the concept of validity to them and state D1-D3. That is to say, while surely there are good and bad abductions, they are always deductively invalid: the premises of an abduction do not *guarantee* the conclusion. Thus, the goodness or badness of an abduction is not merely about its form. The content matters. And finally, abductions certainly serve to increase our knowledge. That is, the conclusion (i.e. the explanation) says more about the world than the premises (i.e. the observations).

Now that you are familiar with the three major forms of reasoning to be discussed in this course, let me outline what we will be doing in what follows. First we will study abduction somewhat more closely in Chapter 2. Then we will move on to do the same for inductions in Chapter 3. These chapters are relatively short. Most of the rest of this course will then focus on deductions.

§1.3 Identifying Your Reasoning

What I expect you to have gotten from this chapter is a rough understanding of the three major forms of reasoning, plus some skill in *identifying* them. What I mean by this is I want you to be able to pick a belief from your own set of beliefs and explain whether it is a direct observation or an abduction or an induction or perhaps a deduction. For example, if you look back at an entry in your diary, can you determine the nature of the reasoning behind each sentence? Let us see how this can be done.

Let's say you find this sentence in your diary: "Today I talked to Mom on the phone". Well, this is an observation. You saw and heard yourself talk to your mom, so you didn't have to *infer* that you did so. The sentence "The cookies were not very delicious" is also an observation, if you tasted the cookies yourself (so you see that when we say "observation" we don't necessarily mean seeing; it can be any of the five senses or vivid feelings such as anger, fear, etc.). But suppose you said "The cookies were probably not very delicious because Uncle Bob loves cookies and when they are delicious he eats them all in two minutes". In this case, that the cookies were not delicious is not something you observed first-hand but something you inferred based on your observations. The reasoning in the above sentences is more or less explicit:

19) When the cookies are delicious, Uncle Bob eats them in two minutes.

20) Uncle Bob didn't eat the cookies in two minutes.

---

21) Therefore, the cookies were not delicious.

This is a valid deduction and has the form:

19') If P then Q

20') Not Q

21') Therefore, not P

This is also a very famous form which we will talk about in this course. You might also want to know what justifies each of the premises. 20 is clearly an observation. It was not explicitly mentioned in our hypothetical diary above, but it is implied (it is a "hidden assumption"). 19, a sentence directly from the diary, is *not* an observation. In general, you cannot *observe* a conditional claim, i.e. a claim with the "if … then …" structure. 19 is an induction. If you are justified in believing this about your uncle it is probably because you have seen him do this several times in the past. Note that this is what the sentence itself implies. A slightly different set of sentences could make things very different. For example, suppose the sentences are "Uncle Bob ate one cookie and didn't touch the rest. The cookies are probably not very delicious". This sounds more like an abduction than an induction. It seems that the person thinks the best explanation why Uncle Bob didn't eat more than one cookie is that the cookies weren't delicious. This may or may not be the best explanation depending on the rest of the person's observations and the alternative explanations on the table. For example, if one of the observations is that Uncle Bob is usually very worried about his calorie intake, then the best explanation need not be that the cookies didn't taste good. But that does not concern us here, because we are not asking whether the person's (or our own) reasoning is good. We are only interested in knowing what made the person say what they said. In this case it is an abduction (whether a good or a bad one).

As our last example, consider these sentences. "I broke two China cups. Mom is going to get really mad!" The first sentence is an observation. What is the second one? It is a sentence about something in the future, so it cannot possibly be an observation. Also, as a general rule, claims about the future cannot be abductions either. The reason is that an abduction is an inference to the best explanation of *what has happened* and is therefore always about the past or the present. So we are left with induction and deduction. The conclusion of an induction as we defined it above is a generalization. But our sentence "Mom is going to get really mad" is not a generalized claim. It would have been if it said "Mom always gets really mad when …". So, our sentence must be the result of a deduction. But one can see how there is also an induction involved. First we make the following induction:

I broke something $time_1$ and Mom was really mad.

I broke something $time_2$ and Mom was really mad.

Etc.

Therefore, Mom always gets mad when I break something.

Then, we use the conclusion of this induction in the following reasoning:

Mom always gets mad when I break something.

I broke something this time.

Therefore, Mom is going to get really mad.

The above examples should help you figure out most statements that can be encountered in everyday life. Here are some rules of thumb for these "diary" exercises:

- Propositions about the future cannot be observations or abductions.
- Propositions involving words like "probably", "must", and "have to" cannot be observations (why?).
- Conditional propositions ("if … then …") are not observations.
- Propositions involving "always", "never" and the like are *likely* to be inductions.
- Propositions involving "it seems that" and the like are *probably* abductions (why?).
- The word "because" is a reliable sign of abductions. The sentence that comes before "because" is usually an observation and the one after it usually an abduction.
- If the statement is not a generalization, it cannot be an induction, even if it is *based* on an induction. For example, "Mom is going to get angry" is based on past experience, but as stated the sentence is just about the singular event of her getting angry *this time*.
- When you decide that something is the result of a deduction, you need to ask yourself what kind of reasoning is behind each of the premises. It might be observation, induction, abduction, or some other deduction.
- Remember that the sentence in question has to figure as the *exact conclusion* of the argument. For example, suppose the sentence is "I was walking to school". Sometimes students respond to this by saying "this is an induction, because I have walked to school everyday in the past so I know I will walk to school today". But the sentence in question is not "I will walk to school today". The exact sentence is "I was walking to school". Even if it were "I will walk to school today" it couldn't have been an induction as we defined it, because it is not a generalization. The sentence "I walk to school everyday" could potentially be an induction, but that is not the sentence in the diary either. You have to work with the exact sentence and you have to place it as the conclusion. So, make a guess for what the argument is and ask yourself: could this exact sentence be the conclusion of this argument? In this case, there is no argument at all, because you can

clearly see and feel yourself walking to school; you don't have to make any kind of inference for it. It is an observation.

- When we ask what justifies a belief, we mean *at the time* the belief was formed. You have to remember the situation at that time. The question is not about what could justify that belief now that you are looking back at it. So, "I was walking to school" is no longer an observation, but it *was* at the time you formed the belief.

- Keep in mind that the answer is not always unique. It might be possible to formulate your reasoning in different ways. It might help to introspect to see how *you* came to believe that sentence. But even with introspection, different answers might be equally good.

§1.E Homework

HW1

1- (1.5 pt.) Explain each of the three major types of reasoning discussed above (each in three or four sentences). Use your own understanding, don't parrot the reading.

2- (0.5 pt.) Consider this deduction: "When he doesn't call me back, he's probably mad at me. He didn't call me back. Therefore, he's probably mad at me". The conclusion of this deduction does not express 100% percent certainty (it says "probably"). But in the text, I said that deductions provide us with 100% certainty. In what sense do we have 100% certainty in this deduction? (Your answer need not be longer than 2 sentences.)

3- (0.5 pt.) Deductions are 100% certain and indisputable. But while scientists often use deductions, they also frequently use inductions and abductions. Why don't they stick to deductions? What is the drawback to deductions? (Your answer need not be longer than 2 sentences.)

4- (0.5 pt. each) For each of the following sentences, write down a valid deduction that has the given sentence as its *conclusion*. Explain what type of reasoning justifies each of the premises.

   a. "Heather is going to the movies tonight."

   b. "Someone has been eating my cookies."

5- (0.5 pt. each) Give an example for each of the following or if it is impossible, explain why.

   a. An invalid argument with all true premises and a true conclusion.

   b. An invalid argument with all true premises and a false conclusion.

   c. A valid but unsound argument with a true conclusion.

   d. A valid argument with all true premises and a false conclusion.

6- (0.5 pt. each) Briefly explain why each of the following statements is true.

a. "I must get up now or else I'll be late" is not the result of an observation.

b. "I could barely stay awake during class" is not the result of an induction.

c. "John will pick up the kids from school today" is not the result of observation, abduction, or induction and therefore has to be a deduction.

7- (3 pts.) Pick 6 sentences from a recent entry of your diary (if you don't keep a diary, write down 6 sentences that you uttered/thought of today). For each sentence identify whether it is an observation, an induction, an abduction, or a deduction. Treat each sentence individually. If the sentence is an abduction, list the relevant observations and some alternative explanations. If you decide that it is an induction or a deduction, write down the premises and the conclusion. Make sure there is some variety and they are not all observations. Example:

Entry: "The chemistry class was boring today, even though I usually enjoy chemistry. I guess the professor was tired today. I think I'll have a hard time doing the homework tonight."

"The chemistry class was boring": Observation

"I usually enjoy chemistry": Induction

Premises: I enjoyed chemistry the first class period last semester. I enjoyed chemistry the second class period last semester. Etc. (For most of last semester). Conclusion: I usually enjoy chemistry.

"The professor was tired today": Abduction

Observations: The class was boring. The professor spoke slowly. He couldn't focus like before. I talked to a few other students and they agreed with me that class was boring.

Possible explanations: E1: The professor was tired. E2: The professor has suddenly lost interest in the subject. E3: It's just me; the class wasn't actually boring.

"I'll have a hard time doing the homework tonight": Deduction

Premise 1: When I don't like the class, I have a hard time doing the homework. Premise 2: I didn't like the class. Conclusion: I'll have a hard time doing the homework tonight.

Further Reading

Godfrey-Smith, Pether (2003), *Theory and Reality*, University of Chicago Press, Section 3.2

Copi, Irving M., Cohen, Carl, and McMahon, Kenneth (2011), *Introduction to Logic*, Pearson Education, Inc., Chapter 1

Chapter 2

Abductive Reasoning

§2.1 How to Set up An Abduction

In Chapter 1 we learned roughly what an abduction was and how often it is used both in everyday life and in science. In this chapter we will study this form of reasoning in much more detail and learn how to improve our abductive skills. To set up an abduction, the first thing to do is carefully distinguish what we are observing from what we are not. This might sound simple, but it can be a formidable task because people tend to jump to conclusions about what is happening. In other words, people tend to confuse what is in fact merely a hypothesis with what they are actually observing. I made the example of another person's hatred on p. 12. You don't (and you can't) actually see or hear or smell or touch or taste hatred. So that someone hates you is never an observation. It is always the *explanation* for what you see, hear, etc. This is not specific to hatred. Love, anger, and most other emotions are not observable. Even our own emotions are often confused or subconscious. We just guess what they are.

But this mistake does not only occur about emotions. People often say things like this: "I was really sleepy this morning, but the first cup of coffee woke me up". This is said in a way that makes it sound like two observations; one to the effect that the person was tired and the other to the effect that the coffee woke him/her up. However, we don't observe the workings of the coffee here. You know you were sleepy, and you know the drowsiness went away shortly after the first cup of coffee. These are the two things you actually observed. But since you think that the *best explanation* for the drowsiness going away is that it was the coffee's doing, and since that seems very obvious to your mind, you might think you almost observed it. In regular situations this sloppiness might be OK, but when it gets tricky, it is very important to distinguish your observations from non-observations. Consider this example: Someone gives you what they claim is a potion to relieve your pain. You drink the potion and soon the pain goes away. You then think the potion was effective, just like the coffee. But in fact closer scrutiny reveals that the potion was nothing but simple tap water. So there has to be another explanation why the pain went away. A lot of times it is just the reassurance and feeling of safety that the placebo gives you that takes away some of the pain. This is how many people fall for certain ineffective types of alternative medicine. You might have heard some of those people say: "I have *seen* it work", which is another example of conflating explanations for observations. This example shows that even in the case of coffee, the explanation remains an explanation and should not be elevated to the status of an observation.

So, the first thing to do is list all your observations, being very careful not to build any bias into them. This is what we did in the last chapter, in the case of footprints on the beach:

O1) There are footprints on the sand.

O2) The footprints have the shape of human shoes.

O3) The footprints are regular.

O4) Everything else seems to be normal.

Note that there might be no principled end to the list of your observations, but we usually have a sense of where to stop. If other observations prove relevant later on, we will have to add them to the list. The next step is to list some of the best explanations that you can think of. Be careful that no matter how obvious one of the explanations might seem (and usually one of them does seem obvious), you cannot simply assume that it is true until you go through the abductive process. You have to learn to be skeptical about things that you take for granted. Here are some possible explanations for O1-O4 from the last chapter:

E1) Someone has been walking here.

E2) A monkey with human shoes has been walking here.

E3) Aliens have created these marks with some high technology to deceive me.

E4) Through a rare but possible process, the wind has accidentally created these marks on the sand.

E5) Some invisible ghost has made the marks on the sand. The ghost is extremely careful not to be seen or sensed and will therefore never create any hint while you are watching, but it does things when no one is paying attention.

The list of possible explanations is also technically endless, so we have no choice but to work with a few that we can think of. As I said, E1 might seem obvious to you and the other explanations might seem crazy. But we want to know exactly *why* it is that E1 is the best explanation. What features single out E1 as the best? Philosophers have come up with a number of such features. We will study only four of them in this course. So, in the next four sections, our goal is to find out four of the most important features in which E1 scores better than E2-E5.

§2.2 Simplicity

Perhaps the single most important feature of best explanations is that they are the simplest explanations. Simplicity has rather precise definitions in logic. But let me first tell you what it is *not*. Simplicity has nothing to do with ease of understanding. So, a simple explanation might be very difficult to grasp, while a not-so-simple one might be easier to understand. Most physicists think that Einstein's theory of General Relativity is the simplest explanation of large-scale physical phenomena. Yet most physicists agree that General Relativity is an extremely confusing and cumbersome theory both in terms of comprehending its concepts and in terms of carrying out its calculations. How long or short an explanation is also has nothing to do with how simple it is. An explanation can take more words to express in our language but it might be simpler.

So what is simplicity? There are different definitions out there. In this course, we will work with two similar but slightly different senses of simplicity.

> *Sense 1 of simplicity: The simpler explanation is the one that needs fewer entities in the picture, i.e. populates the world by a smaller number of things.*
>
> *Sense 2 of simplicity: The simpler explanation is the one that needs to make fewer assumptions in order for the explanation to work.*

Sense 1 of simplicity is sometimes called "Ockham's Razor" after the famous medieval philosopher who argued that we should not assume the existence of any entity unless we need it to explain our observations. Many atheists appeal to Ockham's Razor when asked why they do not believe in God. In their opinion, since everything we know so far about the world can be explained by scientific processes, we don't *need* to assume that God exists. Interestingly, many theists draw on the same principle to argue that God does exist. That is, they claim there are certain facts about the world which can only be explained by the existence of an all-powerful, all-knowing entity. The point is that both sides seem to agree that simplicity is important. We do not want to assume the existence of any entity unless it is necessary for explaining something. That is also why most adults do not believe in unicorns, fairies, demons, etc.

Sense 2 of simplicity needs clarification. Let us assume that Mr. Smith has trouble reading the subtitles when everyone else can easily read them. He also can't drive as easily as before and often asks what the road signs say. And when he sits at his desk, his face is unusually close to the computer screen. These are our observations. Mr. Smith explains the first observation by saying that it is because the subtitles are yellow and he can't stand yellow text, the second one by saying that he has too much on his mind and is distracted, and the third one by claiming that he often looks up tiny pictures. But clearly, the simpler explanation is that his eyes need a new prescription (but perhaps he is so self-conscious about having weak eyes that he does not want to admit it). Mr. Smith's explanation is less simple because he has to tell a separate story about each of the three mentioned observations, while one story can take care of them all. If you can explain everything with just one assumption, there is no need to make three.

In order to correctly count the total number of assumptions, we need to count not only those assumptions that the theory makes explicitly, but also other important questions that the theory needs to answer before it can get off the ground. Sometimes a theory might appear to make very few assumptions, but in fact the theory is such that it raises some further questions that need to be addressed before the theory makes sense. In order to "explain away" those questions, the theory then needs to make more assumptions. Its total number of assumptions might therefore end up being more than its rival. For instance, say I have two observations: 1) I have been eating a lot of fat and 2) the scale shows an increase in my weight. One theory is that I have gained weight due to eating too much fat. This fits both observations and raises no further questions. However, I might prefer another theory that says that the scale is broken to explain why it shows an increase in my weight. This seems like just one assumption at first. But now I need to explain how the scale broke without anything happening to it. I also need to make an assumption about why the extra fat that I've been eating has not affected my weight. This adds up to three assumptions total. Some of these assumptions are not there to address any of the original observations, but rather issues that are immediately raised by the explanation itself. That I have gained weight is a simpler explanation in sense 2. (Note that neither theory assumes extra entities, so they are equally simple in sense 1.)

Sometimes good theories have to explain away certain things. But the more we have to explain away, the worse our theory is. For example, I do not believe in being haunted by ghosts because I have never seen anything that needs to be explained by assuming that ghosts haunt people or houses. However, there are a few stories out there that I need to explain away. I do so by assuming that people who tell such stories are feeding their imaginations and exaggerating certain visual illusions. Some other stories I explain away by assuming that those people are lying or trying to draw attention for psychological or commercial reasons. So my theory (that there is no such thing as being haunted by ghosts) does have a few things to explain away. It would be a simpler theory if it didn't. The theory that there are ghosts and they sometimes haunt people is therefore simpler *in this respect*. But it is less simple in that it assumes the existence of additional entities (namely, ghosts) that are absent from my theory. There are two lessons to draw from this example. First, the best explanations sometimes have to explain away certain things that the alternative explanations might easily accommodate. But they might still be *overall* simpler than the alternatives. Second, simplicity might sometimes not be able to single-handedly decide between rival explanations. We will then have to turn to the other criteria of good explanations (see below).

Let us apply the two senses of simplicity to the footprint example. We can see that E2 assumes the existence of a human to whom the shoes belong *and* a monkey, so it is less simple than E1 in sense 1. E3 also assumes the existence of aliens and is less simple in sense 1. And similarly for E5 and the ghost. E4 does not assume the existence of any additional entity and is therefore as simple as E1 in sense 1. What about sense 2? E1 does not have to explain away any of O1-O4. It does not create any significant new questions either. E2, on the other hand, might have to explain away O3, because normally we do not expect monkeys to walk very regularly. E2 also creates an extra question as to why or how the monkey was there. E3 perfectly fits with O1-O3, but seems a little at odds with O4. If aliens were recently here, how come everything else seems to be normal? But perhaps this is not a big problem for E3. The more important issue would be that it creates a new question to be explained away: Why do aliens want to deceive me in the first place? E4 does not have to explain away any of O1-O4, except perhaps for O3. But since E4 is already assuming that the footprints were created through a rare process, it is not *an extra* problem for this theory to explain O3. We will see below that the rarity of the process is itself a problem though. E5 does not need to explain away any of the original observations, but it does create the additional question of why the ghost would want to leave footprints. Thus, in terms of simplicity, E1 and E4 win.

§2.3 Testability

The next criterion we will consider is testability. The more testable the theory, the better it can serve as an explanation of our observations. But we need to be precise about what we mean by testability. Here is a rough definition:

> *A theory is testable if it gives rise to observable predictions other than the original observations it was built to explain.*

Darwin developed his theory of evolution to explain certain phenomena that he observed regarding the similarities and differences between different species (in particular birds). The theory stated that these species have evolved from a common ancestor through a long, piecemeal process of evolution. It then turned out that if his theory is right, there must have been other species in between, say, humans and chimpanzees that are now extinct. If these species existed, they must have left fossils that we can dig out. This was *not* something that Darwin meant for his theory to imply. It just turned out to be one of the consequences of his hypothesis. Darwin's theory was therefore a very testable theory: if we found those fossils, his theory would be strengthened, and if we didn't, it would be weakened (we have in fact found many such fossils). It also followed from his theory that the Earth must be much older than scientists thought at that time. This was also not one of the original observations that Darwin built his theory for, but a prediction that proved his theory to be testable (this prediction was also confirmed).

Let me make a more concrete example. Suppose I have an accident while driving my car. One explanation for this accident is fate. It was “meant to happen” to me on that day. Another explanation is that my car has a bad brake and needs to be fixed. Both of these theories explain the accident, but the second explanation is clearly more testable than the first, because it tells me something that I can then check for truth (namely, I can check my car’s brake system). The fatalist explanation, on the other hand, has no testable consequences. Of course I might check the brake and find out that nothing is wrong with it, but that is a different issue. As far as testability is concerned, the brake explanation is better than the fate. We sometimes say that fate is an “irrefutable” or “unfalsifiable” explanation. Here is what that means: If my car doesn’t work and I don’t get to my destination, it can be claimed to be fate. If my car does work and I do get to my destination, that can also be fate. There is nothing that could possibly happen which would prove to us that something was *not* fate. People who are not good with abductive reasoning are sometimes proud that their theory explains absolutely everything. But that is always a red flag for unfalsifiability. If your theory is too powerful in explaining stuff, it might be because it is simply not testable, so you can always claim it to be true no matter what happens.

In the case of our mysterious footprints, which explanation is less testable than others? E1-E3 are pretty much on a par as regards testability. All of these theories explain the footprints by postulating the existence of some other entity (a human, a monkey, and alien creatures, respectively), that we can in principle find and observe. E2, for example, has as its consequence that if we follow the footprints, we can hope to see a monkey around. E3 might be a little harder to test because we don’t yet know where to look for the aliens. But that is not a very big problem for E3; some good explanations are rather difficult to test. E4 has some testability issues because the incidence where wind creates footprints is explicitly said to be a very rare phenomenon and thus if we want to test this hypothesis by waiting for another similar occurrence we will not have much of a chance. But E4 might be tested in other ways. For example, someone might argue that if wind has moved sand in this peculiar way, it must have left other signs as well. Leaves in the nearby woods, for instance, must have been shifted around with a recognizable pattern too. Or one might say that for wind to blow in this unusual manner, the barometer over at the weather center must have recorded some quite remarkable activity. (Exercise: Think of other observable tests for E4.) The theory that has a much more serious testability problem is E5. In fact, the biggest problem with this theory that disqualifies it as a good explanation is perhaps its testability. It is built into the statement of the theory that this ghost is totally elusive. It is extremely careful not to be observed or sensed by us. It is also not mentioned what other kinds of things this ghost tends to do. So we have no systematic way of finding this being or any of its possible signs. (This type of lack of testability is often seen in people’s occult beliefs. In certain subcultures there are circles of people who believe in some sort of supernatural power or entity and have no problem sensing it when they get together. But when an outsider expresses skepticism about their beliefs and asks to attend their meetings to observe the supernatural being with them, they respond by claiming that the entity does not like to show itself to people who do not profoundly believe in it. This greatly reduces the testability of the theory, because you are either a non-believer who will never have a chance to observe the entity, or you are a profound believer which is prone to hallucinations about the entity.)

Therefore, in terms of testability, E1 and E2 score the highest, followed by E3 and E4. E5 is in serious trouble.

§2.4 Likelihood

A good explanation should also be likely to happen. Other things being equal, the more probable theory is the better theory. How do we know what is more likely to happen? Suppose that you have been roommates with John for about 2 years, which includes 100 Fridays. During this time, John has come home drunk on about 60 Fridays. So you know that it is almost 60% likely that John comes home drunk on a Friday. He has only been drunk on 5 Saturdays, however, which means that it is much less likely (about 5%) that he comes home drunk on a Saturday. This method of assessing the likelihood of a given scenario is based on the past frequency of its occurrence. Sometimes we use our knowledge of *similar* events in the past to assess the likelihood of different scenarios. For example, it is more likely for me to see my neighbor than to see my cousin who lives in Middle East on the way to work. If you throw a pair of dice, it is more likely to get two different numbers than to get a pair of matching numbers (like two ones, two twos, etc.). Some of these probabilities can be calculated exactly, while others must be guessed using common sense.

Likelihood is often one of the major problems with conspiracy theories. Suppose I believe in a conspiracy theory according to which all federal government decisions are made by an elite group of 30 people behind close doors while the public is deceived into thinking that there is a democratic process. This theory might be very simple and explain a lot of things, but it is probably not the best explanation of our observations, because it is highly unlikely that all elected officials would cooperate and not say anything, and that no reports would leak out of these meetings to the public. In general, theories that claim that there is a lot going on with regard to something but no one knows about it are unlikely scenarios. It is in particular unlikely that the person proposing the theory happens to be the *only* person knowing about those secret events.

Students sometimes use likelihood in a very broad way to simply reassert that the explanation at hand is not the best explanation. For example, one might say things like “it is more likely that there are no demons" or "aliens are unlikely to exist". These are not correct applications of the likelihood criterion as we defined it. It is not even clear if it makes sense to talk about probabilities in such contexts. If the question is the very existence of demons or aliens, you cannot criticize it by assuming that it is unlikely. We can only use likelihood for things that we know happen. The likelihood is about *how often* it happens. The existence of demons is not something that has several chances of happening or not happening. So we cannot meaningfully talk about probabilities here. (Of course one might belong to a certain branch of a certain religion or belief system that tells them how likely aliens are, but in logic we only draw on *common* knowledge, not this or that person's personal beliefs.) Then why does it sound plausible to say things like "it is more likely that there are no demons than that there are" (or, more simply, "most probably there are no demons")? When people say things like that what they mean is that they don't think it is rational to believe in demons. Their reasons for thinking this might be based on abduction, but this abduction cannot involve the criterion of likelihood.

A nice way of summarizing all this is to say that in this book, by "likelihood" we mean objective or statistical, and not subjective or theoretical, probability. Let me explain. I said that the existence of aliens is not a question of likelihood. But you might have heard of the Drake Equation, an equation that estimates the probability of existence of aliens in the Milky Way galaxy. Isn't that likelihood? It is in one sense, but not in the sense that we are using in this book. The Drake Equation is the *theoretical probability* of existence of aliens, because it is based on a number of theories in biology and cosmology. Likelihood as we defined above, however, is the statistical frequency of the phenomenon, and the existence of something does not have a frequency. It either exists or not. It does not make sense to ask *how often* something exists.

Among our footprints theories, E2 and E4 have likelihood issues. A monkey wearing human shoes and walking on the beach is a rather unlikely occurrence when compared to a human walking on the beach. As for E4, the unlikely nature of the event is admitted by the theory itself. The laws of physics allow that wind might shape the sand in this way, but the probability is so low that we practically need to wait billions of years to have a shot at witnessing it. Thus, with respect to likelihood, E2 and E4 score low.

§2.5 Consistency with Previously Well-established Beliefs

This is the last criterion for good explanations that we will discuss. A good explanation should be consistent, as much as possible, with other well-established beliefs such as uncontroversial scientific facts. That is, a good explanation should try to be in tune with everything else that we take for granted. So suppose I find my wallet which I am pretty sure I left on the coffee table to be on top of the refrigerator. One explanation for this observation could be that my wallet defied gravity for a moment and flew to the top of the refrigerator. A rival explanation claims that someone tall enough moved the wallet to where it is now. The first explanation is *inconsistent* with the law of gravity. Thus, we would *not* want to accept this explanation unless we really have to. As long as it is possible to think that a person is responsible for the misplacement of my wallet, we prefer to stick to our firm belief in gravity.

But we must be careful when applying this rule. The criterion of consistency does *not* tell us to stick to any old personal belief simply because it is more comfortable this way. First, the belief in question must be *well-established*, i.e. it has to be something that has been confirmed to be true over and over again in the past. Gravity is one such thing, but that the number 13 brings bad luck is not. So even if *you* firmly believe that the number 13 is unlucky, you cannot use this to argue that an explanation is inferior for being inconsistent with this belief. This belief has not been confirmed using reliable methods (such as scientific methods). So, "well-established" does *not* mean "popular" or "commonly-held". Many commonly-held beliefs turn out to be false or inaccurate on closer scrutiny. Second, the criterion of consistency says that we should not discard our well-established beliefs *unless we have to*. This means that if all other explanations for my wallet's change of position turn out not to work, we might eventually accept that sometimes objects defy gravity. Just like any other scientific theory, gravity is uncertain and subject to correction. But the criterion of consistency asks us to do a great deal of research before giving up gravity.

Which of E1-E5 is problematic with respect to this criterion? E2 might seem to be at odds with some of our beliefs about monkeys and where they can usually be found or what they usually do, but these beliefs do not seem to be very fundamental or unexceptionable. E3, however, runs against scientific consensus on the issue of aliens. As of today, we have compelling evidence that either extraterrestrial intelligent beings do not exist or that if they do they are not anywhere close to our solar system to be able to make contact. While this scientific consensus might in the future turn out to be mistaken, we should be reluctant to accept a theory that goes against it unless it is the best possible explanation of our observations on all other counts. What about E4? It is consistent with all well-established beliefs, other than the fact that it is unlikely to happen, which we discussed above. In other words, E4 is perfectly *possible*, though improbable. E5 is trickier. We might think that there is a well-established scientific consensus about the existence of ghosts, namely that they do not exist. But while it is true that scientists and scientifically-minded people often tend not to believe in ghosts, it is not really the consequence of any well-confirmed scientific theory that ghosts cannot exist. Scientists have not found evidence for ghosts, but they have not found any evidence *against* their existence either. So, strictly speaking, E5 is not inconsistent with other well-established beliefs. Therefore, in terms of our last criterion, E3 is in trouble, while E2 and E5 have unclear fates.

All things considered, E1 seems to win. Thus, logic has confirmed our intuitions about which theory is the best. Note that while we expect logic to confirm many of our intuitions, it is possible that (and often illuminating when) it does not. Sometimes our prejudice or lack of consideration dulls our logical thinking and clear-cut logical rules can help us see through our mistakes.

§2.E Homework

HW2

1- (1 pt.) Explain both senses of simplicity in your own words.

2- (1 pt.) Explain consistency with previously well-established beliefs in your own words.

3- (0.75 pt. each) Which of the theories, if any, is more testable in each of the following pairs? Why?

   a. T1: You didn't get the job because God didn't want for you to get it. T2: You didn't get the job because the employer wanted to hire his cousin.

   b. T1: If you donate to charity, God will never let you get sick. T2: If you donate to charity, the world will be a happier place.

   c. T1: I got poisoned by the soup because these things sometimes happen. T2: I got poisoned by the soup because it had old meat in it.

   d. T1: Capitalist societies are entirely driven by the greed of a few wealthy people and the policies made in these societies are always harmful to the majority of people and never lead to prosperity for the working class. T2: Capitalist societies are entirely driven by the greed of a few people who control the wealth. When the working class is being crushed, there is no reason but the greed of the elite. When the working class prospers, it is also the greed of the elite who are giving the workers just enough money to keep them quiet and avoid strikes.

4- (1 pt.) Suppose your little brother tells you that a squirrel had a full conversation with him. One explanation is that the squirrel did in fact talk to him. One might argue that it is *simpler*, in sense 2, to assume that. Nevertheless, most of us will probably think of another explanation, e.g. that his active little brain is blurring the line between reality and imagination for him, which is a less simple hypothesis (is it?). So, in this case, it appears as though we prefer the less simple explanation over the simpler one. Why is that? Which of the other three criteria compels us to do so?

5- (1 pt.) Suppose you hear this from a fortune-teller: "You might have a lot of problems in your family this year. But if you have a good heart it might not happen to you". What is wrong with the fortune-teller's statement?

6- (0.5 pt. each) For each of the following theories, determine whether the criterion of likelihood (in the *objective sense*) could be used to evaluate the theory at all. Why or why not? If cases where likelihood does apply, say if the theory is likely or unlikely.

   a. The theory that John got rich by winning the lottery 3 years in a row.

   b. The theory that trees have fallen because of storm.

   c. The theory that Santa Claus exists.

   d. The theory that thirty identical-looking unicorns live in Lubbock, TX.

   e. The theory that it is actually not the 21^{st} century and everyone except you knows that.

   f. The theory that there are blackholes.

HW3

1- (1 pt.) Think of a pair of rival explanations for a given set of observations such that one explanation is simpler than the other in sense 1. First list the observations carefully and then list the two explanations. Explain why one is simpler than the other in sense 1. Limit your answer to about one paragraph.

2- (1 pt.) Repeat what you did in the above problem with another set of observations and another pair of rival explanations, this time for sense 2 of simplicity. Explain. Limit your answer to about one paragraph.

3- (1 pt.) Think of a pair of rival explanations for a given set of observations such that one is more likely than the other, but the more likely theory is less testable. Explain.

4- (7 pts.) Answer the following question in one or two page(s).

How do you know you are not dreaming right now? The skeptic's answer to that question is "I don't know". But many people think there are good reasons for you to think that you are not dreaming and those reasons can be put in terms of abduction (inference to the best explanation). On the flip side of that, we can ask how we know dreams aren't real. This is a simpler question, so let's try it. Consider the following observations:

1- In what we normally call "dream", when people fall from high buildings, sometimes they die and sometimes they don't.

2- In what we normally call "dream" sometimes people can fly and sometimes they can't.

3- In what we normally call "dream" sometimes animals can talk and sometimes they can't.

4- When we go from the “dream world” to the “real world” we usually find ourselves in the same place we last were before going to the “dream”, but when we go from the “real world” to the “dream world” we don’t usually find ourselves in the same place as the last “dream”.

Now consider these two rival theories:

T1: The “real world” is the only actual world and the “dream world” is unreal and the result of some kind of random mental activity.

T2: The “dream world” is also real (like a parallel universe). Observations 1-3 each have their own explanation in terms of natural laws governing the “dream world”. Observation 4 is just a trick that the “real world” plays on us, by transferring us to the same place every time we are about to go to that world.

Is T1 the better explanation? Make sure you address the following:

a- Is T1 the simpler theory? Why or why not? Consider both senses of simplicity.

b- Is T1 more testable than T2? Why or why not?

c- What other observations do you think would be relevant to judging between T1 and T2? How does each theory rank with regard to explaining those additional observations?

d- Are the criteria of “likelihood” and “consistency with previously well-established beliefs” relevant here? Why or why not?

Further Reading

Copi, Irving M., Cohen, Carl, and McMahon, Kenneth (2011), *Introduction to Logic*, Pearson Education, Inc., Chapter 13

Ladyman, James (2002), *Understanding Philosophy of Science*, Routledge, Chapter 7

Chapter 3

Inductive Reasoning

§3.1 Simple and Sophisticated Inductions

There has been (and there still is) a great deal of discussion about inductive reasoning in philosophy and logic. Some of these discussions are concerned with whether induction is ever justified at all. Does the fact that the Sun has risen everyday in the past have anything to do with what it will do tomorrow? Do we have any reason to generalize our observations or expect the future to be like the past? There have been various answers to this (very interesting) question. We will not concern ourselves with this question in this book. From now on, we will assume that induction is justified if it is done well and merely inquire about how it can be done well. This is the other part of those discussions. John Stuart Mill, for example, is famous for formulating five ways in which inductions can be done. Although I will not discuss Mill's methods, much of what will be said in the following can be inferred from Mill's ideas. We shall take up a more critical approach, in which we learn how to conduct a good induction by learning how to criticize a given inductive study. But first let us see what kinds of inductions there are.

The simplest kind of induction is what we saw in Chapter 1. It is simply an inference from several instances in which something has occurred to the conclusion that the same thing will occur in the future under the same circumstances. The philosophers' favorite example is:

Swan$_1$ was observed at time$_1$ and place$_1$ to be white. }

Swan$_2$ was observed at time$_2$ and place$_2$ to be white. } The sample

Etc. }

Therefore, all swans are white at all times and places.

As you can see (and as I said before) the premises do not guarantee the conclusion (unlike a deduction), but they *support* it. That means that if you have seen enough swans you have pretty good reason to be confident in the conclusion. The set of swans that you observed are called your *sample*. So, induction is based on inferring a general conclusion about the entire population based on what is true of a (rather small) sample.

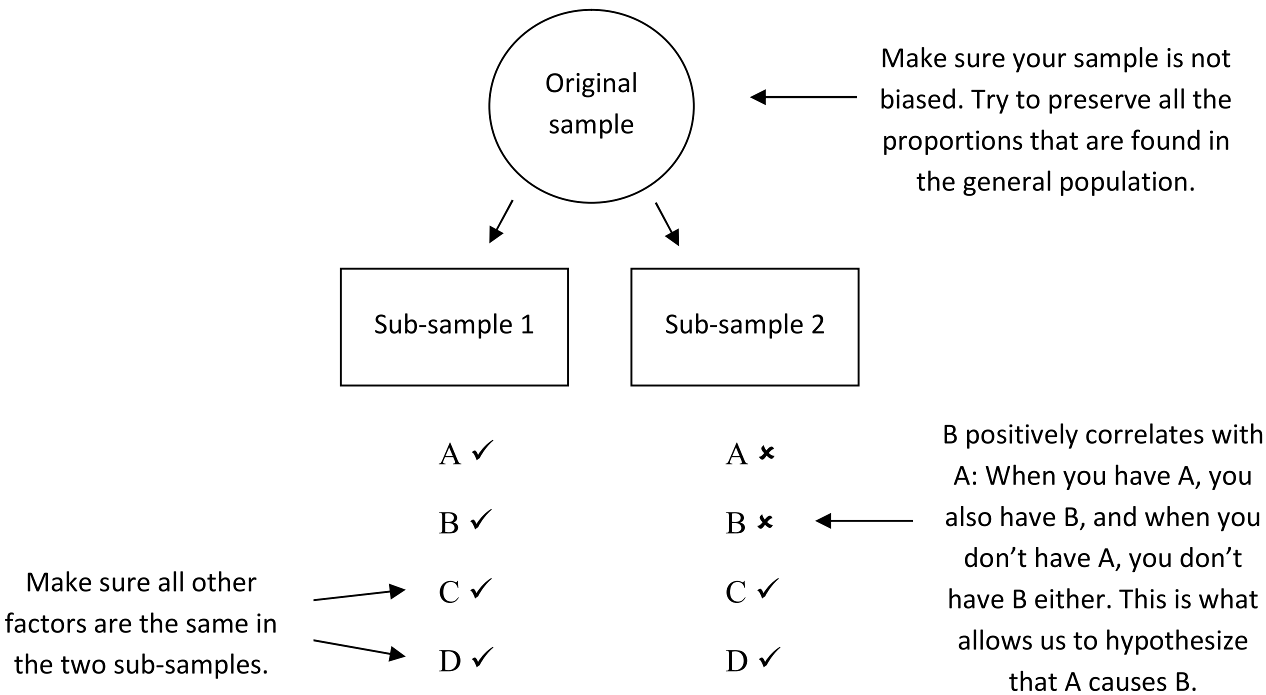
We use this kind of reasoning quite often and it is quite impossible to live without it. However, there are also other, more sophisticated forms of induction. Especially in science, inductions are done by *controlled sampling*, which is a form of reasoning indispensible across the board, in physics, chemistry, genetics, biology, psychology, sociology, and political science. The first step in such studies is to make sure your sample is large and varied enough. For example, if you are studying the American people, you need to make sure that young and old, Democrat, Republican, and Independent, and people from all across the country and from all races are represented in your sample. Otherwise we say that your sample is *biased*. If there are certain natural proportions in the population, we would like our sample to have the same proportions. For instance, if African Americans constitute 13% of the population, we want about 13% of our sample to be African American.

The next step is to divide this sample into two (or more) groups, such that *on average* the members of the two groups don't differ in any feature *except for one*. Here is what we mean by not differing on average: Suppose you divide your sample into two groups and both groups have people from the age of 18 to 78, but the average age of both groups is 40. This means that you did a good job splitting your sample into two. Even though different individuals differ in their ages, the *average* ages of the two sub-samples are the same. More accurately, the *distribution* of age among the members must be the same in the two groups. Ideally, we want this to happen to all factors (age, race, etc.) except for one factor that we are testing. For instance, if you are interested in the differences in the level of subjective happiness of philosophers and non-philosophers, you should split the sample into two groups such that each group exhibits the same distribution of age, income, political allegiance, geographical origin, race, etc., i.e. such that the averages/distributions of all these factors are equal in the two groups. Since the factor whose effects you want to study is being a philosopher, one group must consist of philosophers while the other consists of non-philosophers. This way you have *controlled* for all factors: Since your two subgroups are the same with respect to all factors other than interest in philosophy, you are sure that whatever difference you discover is not due to any of those other factors. Since your two samples are *not* the same with respect to interest in philosophy, the discovered difference, if any, can be attributed to the difference in that factor (see the diagram on the next page).

Once we've gone through these steps, we can then make the following inductive move: we have seen several cases of philosophy causing a difference in people's level of happiness in such and such a manner; therefore, in general philosophy causes difference in happiness in such and such a manner. Let's develop the example more accurately:

> A sample of 1008 American students including all ages, incomes, political allegiances, geographical origins, and races were divided into groups of philosophy majors and non-philosophy majors. Both groups were asked to rank their own subjective happiness in a scale of 1 to 10. Results: 47% of philosophers and 68% of non-philosophers reported numbers higher than 5. Thus, philosophy reduces the overall happiness of the people who study it.

(I don't know if anyone has done such a study or not; I made it up!) This is a pretty strong induction. It is not easy to dismiss the conclusion. But inductions are always prone to mistakes, and a lot of times through no fault of the researchers. For one thing, it is never possible to control for *all* factors that could possibly contribute to the result. Sometimes we don't know that a certain factor also contributes or that we failed to control for it. Below, we will learn four specific ways to criticize and improve an inductive argument such as the one given above. One of these criticisms regards the size of the sample while the other three have to do with controlling for other contributing factors.



§3.2 Sample Too Small

The first thing to worry about when presented with an induction is whether the sample is large enough. Observing a thousand white swans provides a much stronger support for the conclusion that all swans are white than observing two white swans. The reason is that those two swans might have happened to be white by accident. There might be no particular reason behind this accident whatsoever. As the sample grows larger, however, such accidents become unlikely. In the case of sophisticated inductions, the issue of sample size is a little more complicated. But the idea is the same. You can't conclude that all White people are racist because you saw two White people who were racists and two non-White people who were not. No matter how varied and randomized, two samples of two people each is not enough. It might be an accident that the members of one sample were all racists.

But we must be careful in applying this criticism. It is tempting to say things like this: “There are 300 million Americans. How could you make such a generalization based on 1008 responders?”. But in fact 1008 people make for a pretty good sample size for a study such as the one above (the Unhappy Philosophers). The reason is this. The size of the sample should not be compared to the size of the general population, but to the degree of *variety* within the general population. If every American had a truly unique opinion about a certain issue, for example, no sample smaller than the entire population would be large enough. If, on the other hand, all Americans had the exact same opinion about that issue, you could just ask one person and generalize the answer to everyone. In reality, the degree of variation is somewhere between these two extremes. We need to adjust the size of our sample depending on how much variety we have. Of course we might not know the exact degree of variety beforehand (otherwise we wouldn’t need to conduct a study), but we have to use our understanding of how diverse the population is.

Mathematically, there is a measure called *standard deviation* which tells you how much fluctuation you had in your data. This is the number that you find in front of the ± sign in research papers. So, for example, suppose the study reports the number of non-philosophers with a level of happiness higher than 5 (in a scale of 1 to 10) to be 64% ± 3%. That means that most samples of the same size would result in a number between 64 – 3 and 64 + 3. That’s not too big of a difference, so your results would be meaningful. If your standard deviation is too large, however, you need to enlarge your sample to make your results more reliable.

To further clarify the issue of sample size, consider these two inductive studies:

> Study 1: A group of 60 people was divided into two groups of 30. One group tried the new shampoo (claimed to reduce hair loss) while the other group used regular shampoo. No significant difference was found in the levels of hair loss in the two groups. Therefore, the new shampoo is not effective.
>
> Study 2: A group of 60 people was divided into Christians (30) and non-Christians (30). No significant difference was found in the two groups’ opinions regarding afterlife. Therefore, Christians and non-Christians have the same ideas about afterlife.

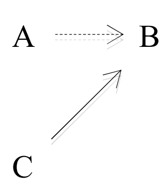
30 people may not be too small of a sample for the purpose of the first study, but it might be too small for the second, while both studies generalize their conclusion from 60 people to the entire population of the world. The reason is that there is a lot more variety in religious belief than there is in physiology and types of hair. Put differently, it is much less likely that the result of the first study is a coincidence than the second. Notice that this is true even if the people in both studies are chosen from across the world, with different ages, races, and so on. Mathematically, we expect the standard deviation of the second study to be a larger number than that of the first study.

When deciding whether a given induction's sample is too small or not, if you are given the standard deviation, use that to inform your judgment. If no standard deviation is given, use your own understanding of the amount of variance you expect in the quantity or quality in question.

§3.3 Uncontrolled Factor

Even when we do our best to make sure other relevant factors are controlled for (as scientists usually do) it is possible that some factors are left uncontrolled. A lot of times this is because the sample is selected in a way that some sort of bias is built into it (often without us knowing). To have an example that does not involve people (remember that inductions are done in all kinds of sciences, not just social sciences), consider a biologist who takes a large sample of bacteria and transfers half of them into another container to study the effect of temperature on them. So he has two groups of bacteria that have the same characteristics on average. Now he keeps one of the containers cold and the other at the room temperature. Let's say he finds out that the bacteria in the colder environment reproduced at a lower rate. So he concludes that low temperature reduces the reproductive activities of the bacteria. Later, his assistant discovers that the container he used for the cold sub-sample was contaminated. The contamination is in this case an *uncontrolled factor* and the results are not reliable, because it might be the contamination, and not the temperature, that caused low reproduction. Regardless of whether or not temperature does in fact have the alleged effect, *this* study fails to show that it does.

As we have seen, a typical inductive study claims that factor A causes factor B because one subgroup had A and one subgroup did not, and the subgroup that had A also had B while the one that did not have A did not have B. An uncontrolled factor criticism amounts to saying that *something other than A* (such as C) might have caused B, because A occurred together with C in all or most of the cases. We can show this diagrammatically as follows:



The uncontrolled factor is C. In the above case, A would be lower temperature, B would be lower reproduction, and C would be contamination. The dashed line means that A was *thought* to be the cause of B, but, according to the criticism, it actually is not.

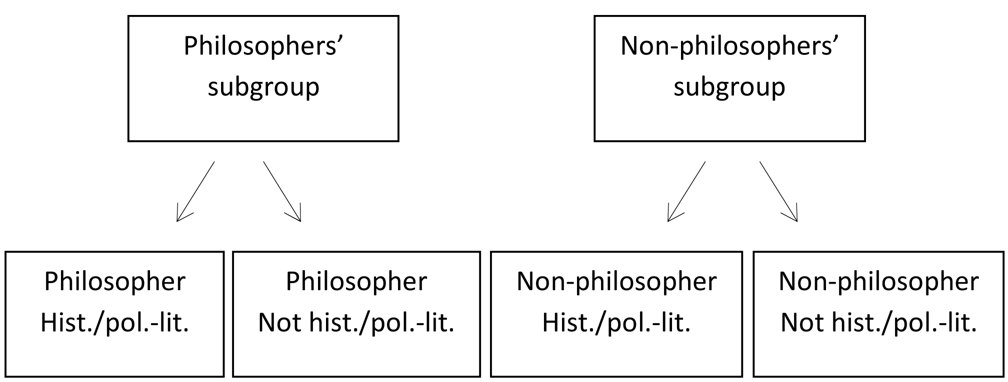
Now consider again the hypothetical case of the effects of philosophy on subjective happiness. Suppose it turned out that most of the philosophy majors were more informed about history and politics than most of non-philosopher majors. Maybe the nature of their major and the typical curricula designed for them require more extensive study of history and politics. In that case, a potential criticism of the study could be that it is not philosophy, but knowledge of history and politics that reduces happiness (say because history and politics involve a good deal of depressing information and make one more conscious about the corruptions of one's society). This would be an uncontrolled factor criticism.

An uncontrolled factor criticism should meet certain criteria before it can be taken seriously. When you come up with a uncontrolled factor, you should ask yourself: What is the relationship between C and A? If the relationship is too loose, then your criticism might be implausible. If the relationship is too intimate, on the other hand, you may not be criticizing the original induction at all. Let's consider these two cases in turn:

Plausibility: In an uncontrolled factor criticism, you are claiming that the factor C *happened to* go together with the factor A, and it was actually C that caused B. So there must be a good story to tell as to why C frequently occurred with A. For example, we had some reason to think that the philosophers in the survey were better informed about history and politics. Our reason was that their major requires deeper study of those subjects both by nature and by design. It was not just sheer coincidence that A and C went together in this case. As another example, consider a typical everyday-life induction. People often make inductive inferences about human beings based on the people around them. For example, they might conclude that most people are unfriendly because most people they have seen were unfriendly. You might want to criticize this induction by arguing that the sample is most likely taken from a certain community where the person spent the majority of his life. So the community would be the uncontrolled factor. This is a good criticism, because there is a strong enough relationship between being in a certain community and being a likely encounter for that particular person. But the following is *not* a good criticism: Maybe everyone he has met is mentally ill. Why should we think that? Until we are given some good reason to take this idea seriously, it is not very probable. It is a *possible* uncontrolled factor, but not a very *plausible* one. It is not clear why this person should happen to run into the mentally ill so often. So it's not enough to just make a guess about some possible uncontrolled factor; one must also provide a story to justify the presence of the factor.

Undermining the original induction: The relationship between A and C should not be too tight, though. Sometimes students come up with "maybes" that they think count as criticism but that in fact do not go against the original induction at all. For example, suppose someone says: "Maybe it's just the nature of people to be selfish and that's why you find them unfriendly". Well, maybe, but that does not really undermine the claim that most people are unfriendly. In fact, it kind of supports the claim and explains it: they are unfriendly because they are selfish. The person who made the original claim about people would not only be unchallenged by this "criticism", but would probably also be helped! A good criticism should show that the person is not entitled to his/her original generalization by showing that something not very closely related to A is responsible for B. This comment does not do that.

Finding a potential uncontrolled factor is not the end of the story. Researchers can deal with such criticisms by conducting further research with improved samples. In the Unhappy Philosophers case, they can make sure that people who know more about history and politics are not overrepresented in their sample. The results may or may not come out the same with the new sample. Sometimes it is also possible to use the *same* sample to respond to the criticism by breaking down the data in a different way. The most thorough response, when your induction is criticized for having an uncontrolled factor, is to divide each of the two subgroups into two further subgroups (so you'll have four sub-samples) in the following manner: Divide the subgroup that has factor A into those who have factor C and those who don't. Divide the subgroup that *doesn't* have A into those with C and those without C as well. Now you have four groups: Group 1: A with C; Group 2: A without C; Group 3: without A with C; and Group 4: without A without C. If the charge of uncontrolled factor is correct (i.e. if C is the sole cause of B), Group 1 and Group 3 will have factor B but Group 2 and Group 4 will not. If, however, the original conclusion is right (i.e. if A is the sole cause of B), Group 1 and Group 2 will have B but Group 3 and Group 4 will not. This way we can decide who is right. You might also get mixed results, if A and C are both partially responsible for some of B. Here is what this procedure looks like in terms of our example above. Take the group of philosophers and divide them up into history- and politics-literate and history- and politics-illiterate. Do the same for the group of non-philosophers (see the diagram below). Now, ask the question: Are history- and politics-literate philosophers less happy than history- and politics-literate *non*-philosophers? Similarly, are history- and politics-*illiterate* philosophers less happy than history- and politics-illiterate non-philosophers? If yes, then philosophy probably is the cause of the difference in happiness. If the answer is no, however, it means that history- and politics-illiterate people, whether philosophers or non-philosophers, show the same level of happiness, and this happiness is greater than that of history- and politics-literate people. In that case, the criticism was right and it is knowledge of history and politics, not philosophy, that causes the difference in happiness.

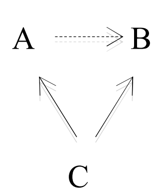


Note that in a case of uncontrolled factor, C does not cause A: being knowledgeable about history and politics does not make you a philosopher! A does not directly cause C either: being a philosophy major does not automatically make you more knowledgeable about history and politics. They just happen to go together in this case (but for understandable reasons). In certain cases, A causes C and then C causes B. We might still call that an uncontrolled factor depending on the case. What about a case where C causes both A and B? Technically, that would also be an uncontrolled factor, but such cases are important enough to deserve their own name.

§3.4 Common Cause

When C is the cause of both A and B, we say the study suffers from a case of *common cause*. Many superstitions are formed because of common causes. Suppose you find yourself thinking of an old friend just before he calls you for the first time since years. You might think that some sort of telepathy occurred, i.e. that your thinking of him *caused* him to think about you. But what in fact happened is you were watching a show on TV that brought up certain memories you had with your friend and he happened to be watching the same show at that time. The TV show caused both of you to think about each other. There was no causal relationship between your thinking of him and his thinking of you.

Here is a more scientific example. It has long been known that the chances of developing many psychological disorders such as schizophrenia are much higher if one's siblings suffer from that disorder. So people who share genes tend to share disorders. This might lead you to think that the genes cause the disorder. The study might be very well done: it has a large enough sample, the sample is divided up into two groups that, on average, don't differ in any respect except siblings of the people in one group have schizophrenia while those of the people in the other group don't. If it turns out that people in the former group have a higher chance of schizophrenia, should we conclude that genes cause schizophrenia? Perhaps. But some critics have disputed this induction by arguing that it might be the upbringing and environment, and not the genes, which causes the disorder. The idea is that people in the same family often experience the same kinds of interactions with their parents and family members, and if those interactions can damage mental health, you might end up with the same disorders as your siblings. So according to this criticism, having the same parents is a factor that causes *both* the similarity in genes and the similarity in mental disorder. This is what we call a case of *common cause*. Here is the diagram for it:



So A would be “having similar genes”, B would be “having similar disorders”, and C would be “having the same parents”. Again, the dashed line means the causal relation between A and B is not real. (What are A, B, and C in the example of the old friend who calls after years?)

Can you think of a common cause criticism for the Unhappy Philosophers case? It has to be something that causes people to study philosophy and at the same time, but independently, causes them to be less happy. One plausible idea would be this: Maybe it is being analytical and the tendency to dig deep that makes these people unhappy, say because it opens their eyes to various uncertainties and dilemmas in life. It is not because they study philosophy, the criticism goes, but because they are generally more analytical people. This tendency to over-analyze things *causes* them to choose philosophy as their major *and* makes them unhappy. If this criticism is right, analytically minded students of other majors are expected to be as unhappy as those in the discipline of philosophy, and not-so-analytical philosophers are expected to be as happy as their counterparts in other disciplines.

Since every common cause is a special case of an uncontrolled factor, the response to a common cause criticism is the same as that of an uncontrolled factor. Divide up each subgroup into two further subgroups, one of which has C while the other does not. In our genes vs. parents case, scientists found a very interesting solution. They compared identical twins to non-identical twins and compared both to other people. This way, you can have four groups: same genes and same parents (identical twins), same genes and different parents (identical twins who grew up apart), different genes and same parents (non-identical twins, non-twin siblings, adopted siblings), different genes and different parents (random people). We then follow the same pattern as explained above in the section on uncontrolled factors to decide whether the criticism was right. If it turns out that people with the same genes, whether they grew up with the same parents or not, are more likely to have the same disorder than people who don’t share genes, the original induction was correct. If, on the other hand, it turns out that people who shared the same parents, whether with the same genes or not, tend to suffer from the same disorder more often than those who had different parents, the *criticism* would be correct and the original induction flawed. (Regarding schizophrenia, both effects were found, so this disorder is currently believed to be most likely the result of a combination of genes and parents.) As an exercise, write down the four groups pertaining to the Unhappy Philosophers induction.

Sometimes the induction is such that there can’t be any common cause. Say, for example, you are given a study that shows young people are on average happier than old people. What would a common cause look like in this case? It must be something that causes both youth and happiness. But nothing causes people to be young other than the fact that they were born on a certain date (unless you believe in youth potions and the like). And the date of birth is likely not a cause of happiness. So it is not easy to come up with a common cause criticism for this case, because nothing *causes* your age in the usual sense of the term that could possibly cause happiness as well. It is very important that you set up a diagram like above and clearly specify what is being claimed to cause what (what is A and what is B). This way you can make sure you don't use the phrase "common cause" haphazardly, because whatever you specify as being your C has to be something that can plausibly be claimed to cause both A and B.

§3.5 Backward Causation

The final case that we consider is when we suspect that it is actually B that causes A and not vice versa. The problem is that when you find a correlation between A and B, you don't immediately know if it is A that gives rise to B or the other way around. Sometimes we know the answer because we know it cannot possibly go backwards. But sometimes it can. There is a funny story about a simple-minded emperor: He heard from his delegates that diseases are more rampant in those provinces where there are more physicians. He then ordered all physicians to be killed because he thought they were causing the diseases! This happens in science too. When studies show that, for example, there is a correlation between hours spent playing video games and aggressive behavior in children, psychologists immediately raise the question "Are video games causing aggression or are aggressive tendencies causing children to be more interested in video games?". This question suggests that what we have here might be a case of *backward causation*. Here is the corresponding diagram:

A ⇄ B

A backward causation criticism for our study of Unhappy Philosophers would argue that, rather than philosophy making people unhappy, people who are already unhappy tend to choose philosophy as their major (say because they think philosophy would provide answers to some of their qualms with life).

Responding to a backward causation criticism is more complicated than the previous kinds of criticism. One way to decide whether or not we are dealing with backward causation is through the so-called *longitudinal or follow-up studies*. We take a randomized sample of children and divide them up into two identical sub-groups. We specifically make sure the two groups exhibit the same levels of aggression. We then let one group play video games and prevent the other from doing so. A few weeks later we compare the two groups again to see if there is any significant difference in their levels of aggression. If the group that played video games turns out to be (even slightly, but statistically significantly) more aggressive, our conjecture has been probably correct that video games cause aggression. If, however, the two groups turn out to show roughly the same levels of aggression, the conjecture has been probably wrong. Follow-up studies are expensive and difficult to carry out. What's worse, they are only possible when the factor of interest is something we can easily control. There are other statistical methods to settle questions of this type, but we will not discuss them here.

Note that in none of the cases we presented in this chapter are we saying that the criticism is in fact right. We are simply listing all possible ways to critique an induction. We also talked about *how* to decide whether the criticism is right. The logician's job is done at that point. It is the scientist's job to go out in the field and find out who is actually right using the logician's methods.

§3.E Homework

HW4

1- (1 pt.) Explain the various steps required for a controlled sampling in your own language.

2- (0.75 pt. each) Using your understanding of the amount of variations in each of the given examples, decide whether a sample of 10 members is too small, rather small, or large enough for each of the following inductions. Briefly explain your answer.

   a. An induction about whether a certain poison is lethal to people.

   b. An induction about whether a certain type of bacteria is resistant to a certain anti-body. (Hint: Bacteria mutate very fast, which is why you need to take your antibiotics on time.)

   c. And induction about people's driving skills in 6 different age brackets.

   d. An induction about people's driving ability after 3 glasses of beer.

   e. An induction about whether china cups break when dropped from a certain height.

3- (0.75 pt. each) An immigrant who has lived and studied in Turkey for five years claims, based on his own experience, that almost all Turkish houses are poorly designed. In each of the following cases a critic has proposed a possibility of uncontrolled factor for this induction. We said above that some uncontrolled factor criticisms should be taken more seriously than others (see p. 36). Which of the following criticisms do you think should be taken seriously? Briefly explain.

   a. Maybe in the Turkish culture, people like the kind of house that he finds to be poorly designed.

   b. Maybe he is attracted to the type of people who tend to like the "bad design".

   c. Since he is a student, he hangs out almost exclusively with students, who can't afford nice houses.

   d. Maybe all houses in this person's sample were designed by the same person.

e. Maybe most houses this person has seen are very old houses (assume that old houses have bad design).

4- (1 pt.) When presented with the possibility of an uncontrolled factor or a common cause, we split up our sample into 4 subgroups to test the correctness of the criticism. Why do we need 4 sub-samples? In what circumstances can we say the criticism was correct and in what circumstances the original induction?

5- (0.5 pt.) Since most of your exercises concern *criticizing* inductions, you might start to think that all inductions are flawed. To counter that effect, think of 2 robust inductions that you believe in.

HW5

1- (0.5 pt.) Consider a study about the relation between race and IQ that shows people of a certain race have lower IQs. Suppose one criticizes the study by arguing that race correlates with socio-economic status and cultural environment, and it is not because of their race that they differ in IQ, but because of their socio-economic status and their culture. Is this an uncontrolled factor or a common cause criticism? Briefly explain your choice.

2- (0.5 pt.) It is generally believed that college graduates are somewhat more likely to be left-leaning and liberal than people who never went to college. In fact scientific studies confirm this commonsensical view. The question, then, is whether this correlation indicates causation, i.e. whether going to college actually makes one more liberal. To answer this question, sociologists Jennings and Stoker followed the political views of 1000 high school students since when they entered college until graduation and monitored the changes in these views. They then compared this change to a similar group of young adults who did not go to college. What kind of study is Jennings and Stoker's? Why did these scientists conduct this study?

3- (0.75 pt. each) Come up with a possible criticism for each of the following cases of inductive reasoning. Identify the type of your criticism and briefly explain it.

a. You acted weird when we went out $time_1$; You acted weird when we went out $time_2$; etc. etc. (four times). Therefore: You are going to act weird when we go out tomorrow.

b. We had a large sample of people with curable diseases and a large sample of healthy people. It turned out that people with curable diseases are more prone to crimes.

c. There is a correlation between money and politics: Lawmakers usually vote for the interests of the lobby that contributes to their campaign. Therefore, money buys the legislators' votes.

d. In a study of a large group of middle-age males and females it was found that people who wear sneakers tend to have healthier hearts. So sneakers decrease your chances of heart disease.

e. I took Logic and it was awesome. Philosophy is great and I wanna major in philosophy now.

f. Countries in Africa are on average much poorer than the rest of the world. You just can't have an African country that's not poor. (Being located in Africa causes poverty.)

4- (0.75 pt. each) For each of the above inductions, explain how one could respond to the criticism that you mentioned. That is, explain how one could decide whether or not the criticism is right by examining more samples, further dividing the sample, conducting a follow-up study, etc.

Further Reading

Gensler, Harry J. (2002), *Introduction to Logic*, Routledge, Chapter 13

Copi, Irving M., Cohen, Carl, and McMahon, Kenneth (2011), *Introduction to Logic*, Pearson Education, Inc., Chapter 12

Chapter 4

Propositional Logic: Symbolism

§4.1 A Simple Model of Language

Now that we know a little about abduction and induction, let us turn to deductions, which is the subject of the rest of this course. We shall focus on modern deductive logic. As you well know by now, deductive logic is all about the *form* of sentences and arguments, not about the content. This is because deductive logic is concerned with the notion of validity (whether the truth of the premises makes it impossible for the conclusion to be false). And as we explained in Chapter 1, valid arguments are valid merely because they follow a correct form. Due to this obsession with form, deductive logic employs symbols to function as place-holders for sentences or parts of sentences. In this way, we can extract the form/structure and forget about the content, as we did to some extent in Chapter 1. So, deductive logic is going to involve a lot of thinking in terms of symbols. You should learn not to be scared of symbols and, if it helps, sometimes fill out the content for yourself (by replacing the symbols with real sentences in your head) to make it more tangible.

We shall study the basics of modern logic in two steps that build on each other. They are called *Propositional Logic* (PL), the subject of this and the next two chapters, and *Quantificational Logic* (QL), the subject of the remaining of the book. Both of these are very simple models of the logical structure of language that nevertheless cover a great variety of deductive arguments. There are many other, more complicated systems of logic that try to improve on PL and QL, but we will not discuss them here.

The smallest units in propositional logic are propositions, i.e. sentences (which is why PL is also sometimes called *Sentential Logic*). So in PL we do not study the internal structure of sentences, but rather the *relations* between sentences. For example, when one utters the sentence "If you break the law, you deserve punishment", one is expressing a conditional relation between two claims, namely "You break the law" and "You deserve punishment". To extract the structure of such a sentence in Chapter 1, we simply replaced each complete sentence with a capital letter and wrote "If P then Q". So "if … then …" is part of the structure of the sentence, not its content. Working with arguments and complicated proofs would be much easier and faster, though, if we replaced "if … then …" with a symbol too. So instead of writing the English words "if" and "then", we can put a symbol in between the two sentences to represent the relation. Some logicians use an arrow that points from P to Q, like this: $P \rightarrow Q$. Others use a "horseshoe", like this: $P \supset Q$. This is basically just an agreement between us about how to represent the form of if-then sentences symbolically. Similarly, instead of writing "P and Q" for a sentence like "Open the door and the window", we can choose a symbol to represent "… and …" structures. Some have decided to use &. Others use other symbols. It does not matter what symbols we choose, what matters is that we build a precise language, free of ambiguities and useful for calculating complicated arguments. Below are the symbolic ingredients I have chosen for our PL language:

Sentence letters: Uppercase letters of the alphabet, such as A, B, P, Q, etc. represent complete and meaningful sentences.

Five Connectives:

NOT: A tilde or squiggle, ~, roughly represents the English "not". We shall call it *negation*.

AND: A dot, **.**, roughly represents the English "and". We shall call it *conjunction*.

OR: A wedge or vee, ∨, roughly represents the English "or". We shall call it *disjunction*.

IF-THEN: A horseshoe, ⊃, roughly represents the English "if … then …". We shall call it *conditional*.

IF AND ONLY IF: A triple bar, ≡, roughly represents the English "if and only if". We shall call it *biconditional*.

Parentheses: We will put parentheses around every *pair* of sentences joined by one of the connectives other than ~.

We will talk about the meaning of each of these symbols below. But first let's see how they are used. We can construct more complicated strings of symbols by putting together simpler ones. But our language has very strict rules for doing this. A string of symbols that cannot be built using these rules is meaningless and incorrect, regardless of how meaningful it might seem in your eyes.

If I write down the capital letter P, I am referring to a complete and meaningful sentence, usually one that cannot be broken down into simpler sentences. So, P can stand for any of "Blackholes exist", "The snow is white", "These cookies are delicious", "All birds can fly", "I was too hungry to wait for the rest of the family to join me", and so on. It is entirely arbitrary what sentence letter we choose to represent our sentence, but once we choose it, we have to stick to the same symbolization until the end of that exercise. In another exercise, it is up to us whether we use that same symbol for the same sentence or not. Also note that the sentence letter must be unique to that sentence. So if we have all of the above sentences in a single exercise, P cannot stand for all of them at the same time. Usually, it makes the exercise easier to follow if we choose a letter that reminds us of the English sentence. So, for example, we might choose B to stand for "Blackholes exist", S to stand for "The snow is white", and so on.

But there are things we are not allowed to represent by a single sentence letter such as P. For example, P (or A or L or whatever) cannot stand for "These cookies are *not* delicious", because that sentence has a "not" in it which we need to make explicit. So if P represents "These cookies are delicious", "These cookies are *not* delicious" would be ~P. To symbolize "These cookies are delicious and the lemonade is great" we need to come up with a sentence letter for "The lemonade is great" and use the AND symbol. If we choose Q as our other sentence letter, the translation will be: (P . Q). Remember that we need to put parentheses around the entire sentence every time we use a connective. Here are some other examples.

I haven't talked to Carol today or I've forgotten the conversation. = (~C ∨ F)

If Bob goes to the party then David won't be there. = (B ⊃ ~D)

North Korea will be a threat to us if and only if we don't negotiate with them and threaten to isolate them. = (K ≡ (~N . T))

If you tell me when and where the party is, I'll be there and I'll act happy. = (T ⊃ (B . A))

The rules of sentence-making are called the *syntax* of PL and the sentences built in accordance with them are called *well-formed formulas*. We can now summarize the syntax of PL.

Syntax of PL:

1- Every capital letter is a well-formed formula.

2- Every sentence constructed by adding ~ in front of a well-formed formula is a well-formed formula.

3- Every sentence constructed by joining two well-formed formulas using any connective other than ~ and putting the result inside parentheses is a well-formed formula.

4- Nothing else is a well-formed formula.

Let's now discuss the meaning of each of the five connectives. In particular, it is very important to know when and why their meanings do *not* exactly match those of their English counterparts.

§4.1.1 NOT

PL is a *two-valued system* of logic. This means that every sentence in PL is either true or false. There is nothing in between. A sentence cannot be neither true nor false or both true and false. Also, a sentence cannot be true for you and false for me. So, for instance, "There are three-eyed aliens in a galaxy 20 million light years away from us" is either true or false. Of course *we* don't know if it is true or not, but it has its "truth value" independently of our knowledge of it, meaning that there is a fact of the matter as to whether those aliens actually exists or not, regardless of whether we are aware of it. Similarly, "Income inequality is harmful to the society" has a truth value: either true or false. It cannot be true for some people and false for others, no matter how different people's opinions are on this issue. Multi-valued systems of logic have been developed in the 20th century, where sentences are allowed to have values other than true or false, but we will not study them here. The connective that we call negation, i.e. ~, *reverses* the truth value of sentences. Since we only have two values (true and false), when P is true, ~P is false, and when P is false, ~P is true. We can show this in a way that is easy to remember by using what is called a "truth table".

P	~P
0	1
1	0

In truth tables, 0 means false and 1 means true. When we have a single sentence such as P, there are two possibilities corresponding to the two truth values. Thus, our table only has two rows. The truth table above shows the effect of ~ for both of these possibilities. It simply summarizes the fact that when P is true ~P is false and vice versa.

The sentence "It is not the case that Americans hate foreigners" is the same as "Americans don't hate foreigners". So the translation for both of them is ~H (assuming we have chosen H to stand for "Americans hate foreigners"). The sentence "It is not the case that Americans don't hate foreigners" must be translated with two negations: ~~H. Even though this sentence means the same thing as "Americans hate foreigners", we are not allowed to translate it simply as H.

The negations must be explicit, because in our logic nothing is assumed to be obvious. We will later *prove* that ~~H is equivalent to H.

Note that the way "not" is used in everyday language is sometimes different from the way our negation works. So, when someone says "I don't have no money", it would be incorrect to translate it as ~~P. The correct translation is ~P, because what they mean is simply that they don't have any money. This particular use of double negation merely *emphasizes* the content of the sentence. *We cannot represent emphases, tones, and the like in our logic.*

§4.1.2 AND

For the conjunction of two sentences, (P . Q), to be true, both sentences must be true. Otherwise, the conjunction is false. For example, for it to be true that "Sara went to the party and James went to the party" both sentences "Sara went to the party" and "James went to the party" must be true. We have demonstrated this in the truth table below.

Notice how the truth table is built. First of all, since we have two sentences, there are four possibilities: P false and Q false, or P false and Q true, or P true and Q false, or both true. You can see that the conjunction is only true when both sentences are true, i.e. the in the last row of the truth table. In general, for n sentences we will have 2^n possibilities. In this course, we will not deal with cases of four or more sentences. So, mainly you'll need to learn truth tables with one sentence (two rows), two sentences (four rows), and three sentences (eight rows).

P	Q	(P . Q)
0	0	0
0	1	0
1	0	0
1	1	1

Conjunction is probably the most straightforward connective that we study in this course. If an English sentence has the word "and" in it, its translation is most likely going to have a "**.**" in it. But there are exceptions to this rule. Compare these two sentences:

Sara and James went to the party.

Sara and James got married.

The first sentence says that two things happened: Sara went to the party, and James went to the party. But the second sentence reports the occurrence of only one thing: Sara and James getting married. Therefore, the first sentence should be translated as (P . Q) while the second sentence should be assigned a single letter such as R.

Another exception happens when "and" is used to mean "if … then …". For example, the surface structure of the sentence "Tell him the truth and he'll start yelling at you" looks like a conjunction, but it is really a conditional: "*If* you tell him the truth, he'll start yelling at you". So this sentence should not be translated with a "**.**" for "and".

§4.1.3 OR

For the disjunction of two sentences, $(P \vee Q)$, to be true, at least one of them has to be true. So, "Either Sara or James went to the party" is true if Sara has gone to the party, if James has gone to the party, or if both have. In other words, $(P \vee Q)$ is only false when both P and Q are false; otherwise, it's true. The truth table will therefore be:

P	Q	$(P \vee Q)$
0	0	0
0	1	1
1	0	1
1	1	1

Compare this with the truth table for conjunction and makes sure the differences are clear to you.

The first thing to notice about our disjunction is that it does not always behave the same way as the English "or". In English (and in many other natural languages) "or" comes with a built-in ambiguity. Sometimes when we connect two sentences with "or" we mean to say that either one of them can happen, *but not both*. Yet there are other cases where "or" is used to mean either one *or both* of the possibilities are allowed. Consider these examples:

(Mother to child:) It's an hour before dinner. You can have either a candy or a cookie.

(Host to guest:) There are plenty of refreshments here. You can have either a candy or a cookie, or … .

In the first sentence, the mother means to say that her child is not allowed to have both the candy and the cookie (because it is almost dinner time). In the second sentence the host is offering all of the refreshments to the guest which means he can have both the candy and the cookie if he wants to. On the surface, the two sentences are virtually identical. It is the *context* that tells us what is meant. This is what we mean when we say that the English "or" is ambiguous. Let's distinguish these two senses of "or". Let's call the first one, i.e. the one that does *not* allow both, "*exclusive or*", and the second one, which does allow both, "*inclusive or*". Now, we can say that

Our disjunction (∨) is defined to always behave like an inclusive or.

This is quite clear from the truth table above. The fourth row of the truth table represents the situation where both sentences are true, and our disjunction is defined to be true in that row. If we wanted to have a connective for exclusive or, we would have to set the truth value of our connective in the fourth row of the truth table to zero. You will see below, however, that by choosing our disjunction to be inclusive, we are not losing anything. We can still represent exclusive or in our logic, but to do that we need to combine at least two of our connectives.

§4.1.4 IF-THEN

Our next connective is the conditional (horseshoe). This connective is somewhat more complicated than the other ones. From the early days of modern logic it was clear that it is especially difficult to represent claims involving "if … then …" clauses in a logically rigorous language. It seemed that either logical clarity and rigor must be sacrificed, or the connective that we end up defining will not adequately reflect the concept of a conditional claim as understood by competent speakers of natural languages. Let me briefly explain why. Suppose I utter the sentence "If you jump out of the window, you'll break your neck". Suppose in reality you do not jump out of the window and (therefore) you do not break your neck. The conditional claim can still be true, if the window is high enough. On the other hand, it might be false, if the window is not very high. So, we can imagine two situations in which "You jump out of the window" and "You break your neck" are both false, but in one situation the conditional claim is true and in the other situation it is false. This means that the truth value of the conditional claim does not seem to be a direct function of the truth values of its components. This is very different from how "and" and "or" behaved. In the case of "and" and "or", knowing the truth value of the components was enough for knowing that of the entire sentence. This anomalous behavior of the conditional connective is bad news, because it prevents us from building a truth table for the conditional. Let me explain.

Suppose we want to build a truth table for the conditional. We might start from the bottom of the table, where things are clearer. If "You jump out of the window" is true and "You break your neck" is true, we probably want to say that "If you jump out of the window, you'll break your neck" was true. This gives us the last row of the truth table.

P	Q	$(P \supset Q)$
0	0	?
0	1	?
1	0	?
1	1	1

Now, consider the second to the last row. Suppose you jump out of the window but don't break your neck. Well, then the conditional claim was definitely false in this case. So we have the second to the last row determined as well.

P	Q	$(P \supset Q)$
0	0	?
0	1	?
1	0	0
1	1	1

Now, it remains to take care of the two remaining rows. How are we to do this? Suppose you do not jump out of the window in the first place. Should we say the claim that if you did you would break your neck was true? Or should we say that it was false? As I said above, on the intuitive level, the answer doesn't seem to have much to do with what you did, but with other conditions in the world, such as how high the window is, whether there is gravity or not, and so on. This is why we said above that the truth value of the conditional does not appear to be a function of the truth values of its components. Logicians say that these conditional claims are not *truth-functional*.

This is a big problem. Propositional Logic is entirely based on truth-functional connectives. That is to say, we really want to have truth tables for all of our connectives. When faced with this problem, logicians decided to introduce a weaker notion of conditional claims, one that is capable of having a truth table. So, logicians decided to go against our intuitions in this case, and assign definite truth values to the first and second rows of the truth table. Once you decide to assign definite truth values to these rows, it is not difficult to guess what they would be. They cannot be both 0's, for example, because then our conditional becomes the same as conjunction (i.e. "and"). Other assignments also reduce our conditional connective to some other connective. The only truth value assignment that gives our horseshoe a unique meaning is the one obtained by setting both rows to 1. And that is what logicians decided to do.

So, after all the struggling, we have our full truth table as shown below. (P ⊃ Q) is only false when the P-claim is true and the Q-claim is false. Otherwise, the conditional claim is true. This truth table in effect says that it can never be that P happens but Q does not, which is part of what every if-then statement says. This "conditional" might not be a perfect representation of how if-then clauses are used in natural languages, but at least it is unambiguous and its truth value is fully determined by the truth values of its components. Other, non-truth-functional conditional connectives have been introduced in logic, which go beyond the scope of our course.

P	Q	(P ⊃ Q)
0	0	1
0	1	1
1	0	0
1	1	1

Since we weakened our concept of conditional claims, the sentence "If you jump out of the window, you'll break your neck" no more means that you'll break your neck *because* you jump out of the window, but simply that when one thing happens, the other follows. So sentences that have "because" in them should not be symbolized with the horseshoe.

If you don't respect her, you don't love her. = (~R ⊃ ~L)

You don't respect her because you don't love her. = Y

So, we treat sentences with "because" as one unit, *not* as consisting of two component sentences. To put it more interestingly, we cannot represent the relation of cause and effect in PL.

The above (rather lengthy) explanation is important for understanding what our conditional connective is and is not. But from now on, virtually all that matters is the truth table given above. Familiarize yourself with it. The P-part of a conditional is called the *antecedent* and the Q-part is called the *consequent*. Here are some quick reminders of the conditional truth table:

When the antecedent is false, the conditional is true, no matter what the consequent.

When the consequent is true, the conditional is true, no matter what the antecedent.

The conditional is only *false when the antecedent is true and the consequent is false.*

The above sentences don't say anything new. They are just different facts about the truth table given above that you could read off the table yourself.

§4.1.5 IF AND ONLY IF

The last connective in PL is the biconditional. As the name suggests, the biconditional is a conditional that goes both ways. So (P ≡ Q) says that P happens when Q happens and Q happens when P happens. In other words, P and Q go together. For example, for "Sara goes to the party if and only if James goes" to be true, Sara and James must be always together: either both in the party or both not in the party. If one shows up and the other does not, the biconditional claim is false.

So, the truth table for the biconditional looks like this:

P	Q	(P ≡ Q)
0	0	1
0	1	0
1	0	0
1	1	1

As you see, the biconditional demands that P and Q have the same truth value. If they don't have the same truth value, the biconditional is false, just as we suggested in the example of Sara and James above.

"If and only if" is not a phrase that you hear a lot in everyday life. It is, however, used frequently in mathematics, philosophical articles, and sometimes in formal speeches. In written texts, "iff" is short for "if and only if". If it helps, you can think of the conditional as a one-way relationship, while the biconditional is a healthy, mutual relationship.

§4.2 Advanced Translations

So far we learned the definitions and basic uses of our five connectives. In this section, we will learn how to translate more complicated sentences and sentences with fancier words. Let us start with fancy words.

§4.2.1 Words Equivalent to "not"

We saw above that "it is not the case that" is another English phrase that gets translated into ~. Phrases such as "it is not true that" should also be treated the same way. All shortened forms of "not" such as "n't" as occurring in "isn't", "didn't", "haven't" and the like must also be translated into ~. Sometimes (but not always) the word "no" functions as "not". For example, "I have no cash" would be ~C, if C stands for "I have cash". But "no" is much more tricky, so we won't be using it much in our examples until we learn Quantificational Logic in Chapter 7.

Consider this sentence: "Jerry is unmarried". There is an implicit negation in this sentence, due to the prefix "un". Some people prefer to translate such sentences with a negation, e.g. ~J, where J stands for "Jerry is married". This might make things too complicated, because there are all kinds of prefixes that indicate some kind of negativity in the adjective they attach to, and they don't always have the same implications. Some adjectives, such as "insane", are created by adding one of these prefixes to some positive adjective (in + sane = not sane), but at the same time they are synonymous with other adjectives, such as "crazy", which do not seem to be the negation of anything else. So, in this course, we will not read negations off prefixes. Treat "unmarried" and "insane" as simple adjectives that attribute positive properties (such as being single or crazy) to objects. Do not translate them with negations.

Rule: "not P" = "it is not the case that P" = "it is not true that P"

Examples:

Jerry is unmarried. = J

I didn't talk to the manager. = ~T

It is not true that Europeans are racist. = ~R

I wasn't sure what to say. = ~S

That guy is insane. = I

It is not the case that the Earth is round. = ~E

One final remark on negations. Suppose you choose B as the translation for "Bob has a PhD". The sentence "It is not true that Bob has a PhD" would then be ~B. But the sentence "I don't think Bob has a PhD" definitely must *not* be translated as ~B. When you say "I don't think P", you're not denying that P. You're simply denying that *you think* P. (While you might not think that Bob has a PhD, it might nevertheless be true that he has one.) So, the correct translation of "I don't think Bob has a PhD" is not ~B (because B has been chosen to stand for "Bob has a PhD"), but it *is* the negation of "I think Bob has a PhD". So we are allowed to translate it as, say, ~T, where T stands for "I think Bob has a PhD".

§4.2.2 Words Equivalent to "and"

What is the content of the sentence "John got up early but he missed the bus"? At the very least, the sentence is saying that two things happened: John got up early and John missed the bus. But it seems that the sentence is also saying something more. The speaker is apparently somewhat surprised that both of these things happened, which is why they use the word "but". When we use "but", we usually mean to express either surprise or some sort of contrast between the two things that happened. This emotional response to the content of the sentence is not important for the purposes of Propositional Logic. As was mentioned above, we cannot represent things like tones and emphases in PL. Add emotional reactions and surprises to this list. For all purposes of PL, "but" is the same as "and":

John got up early but he missed the bus. = (J . M)

There are several other words that function similarly. "Although", "however", "nevertheless", "nonetheless", "moreover", "furthermore", and other similar terms are all similar in that they say "and" while adding some further emotional or tonal tinge to it.

Rule: "and" = "but" = "although" = "however" = "nevertheless" = "nonetheless" = "moreover" = "furthermore"

Examples:

Although I tried to be nice to her, she was very mean to me. = (T • M)

His papers are very well-written. However, I don't agree with his arguments. = (W • ~D)

We explained it very clearly. Nevertheless, they looked quite confused. = (E • C)

I didn't make it clear that I didn't want to see him again. Nonetheless, I returned all his stuff. = (~M • R)

You should be able to make your own examples for "moreover" and "furthermore".

§4.2.3 Words Equivalent to "or"

We said above that our disjunctions represents inclusive or. So when we say (P ∨ Q), we mean at least one of P or Q is true, but we allow that both be true as well. So, our disjunction literally means "at least one". For this reason, if a sentence starts with "at least one", we can usually translate it with ∨.

Rule: "P or Q" = "at least one of P or Q"

Here is an example:

At least one of the French, the Dutch, or the Greek will win. = $((F \vee D) \vee G)$

(Note the manner in which the parentheses are added.)

§4.2.4 Words Equivalent to "if… then…"

First of all, notice that "If you want it, you can have that TV" is the same as "You can have that TV if you want it" and both must be translated into $(W \supset H)$. Don't get confused by the order in which the sentences come. It does not matter that the sentences are switched around: "you want it" is still what "you can have that TV" is conditioned upon. So remember the following rule:

The sentence that comes immediately after "if" is always the antecedent, regardless of the order of the sentences.

Now we are ready to learn some more fancy words. There are several English phrases that mean the same thing as "if… then…". For example, "We are going to protect you, *provided* you get the work done" just means "We are going to protect you if you get the work done", which is the same as "If you get the work done, we are going to protect you". Words such as "assuming that", "given", and "in case" function similarly.

Rule: "if" = "in case" = "provided (that)" = "assuming (that)" = "given (that)"

Examples:

We are going to protect you, provided you get the work done. = $(W \supset P)$

Assuming that we make these cuts, the deficit will be reduced. = $(C \supset R)$

We need to raise taxes on the rich, given our enormous debt. = $(D \supset T)$

Give me a call, in case I forget to call you. = $(F \supset C)$

Notice carefully the order of the sentences in the original English and in the logical translation. Remember again that *it does not matter* which sentence comes first in the English version, but which sentence comes after "if" or the phrase equivalent to "if".

Another fancy way of saying "If P then Q" is to say "P only if Q". "Only if" works in the opposite direction of "if". That is, what comes after "only if" is always the *consequent*. "Only if" might come in the middle or at the beginning of the sentence. The advice is again not to get confused by the order of the English sentences. It does not matter which sentence comes first. What matters is which sentence comes after the connective. "Only if" phrases are in my experience one of the most confusing things for logic students. So make sure you understand this point very clearly.

The sentence that comes after "only if" is the consequent, not the antecedent.

There is a tendency to confuse "only if" either with "if" or with "if and only if". In reality, it behaves differently from both.

"Only if" is not *the same as "if".*

"Only if" is not *the same as "if and only if".*

Since "only if" is not a phrase that you hear a lot in everyday life, you must understand its behavior by memorizing at least some of the above rules of thumb.

In philosophy and mathematics you often encounter sentences of the form "A is a necessary condition for B" and "A is a sufficient condition for B". These are conditional claims too. One is equivalent to "If A then B" and the other is equivalent to "If B then A". The trick is to learn which goes which way. When we say "A is sufficient for B", we are saying that if A occurs, B is guaranteed to occur, because occurrence of A is enough (sufficient) for the occurrence of B. Thus, "A is a sufficient condition for B" means "If A then B". The other one, namely "A is a necessary condition for B" is the converse: "If B then A". Here is an example that helps clear this up. Water is necessary for making coffee. So if you have a cup of coffee, you know for a fact that you have water (among other things). Hence "If coffee, then water". It might be more intuitive to say "If no water, then no coffee". The latter sentence is called the *contrapositive* of "If coffee, then water", and is logically equivalent to it. Sentences with "necessary" sometimes sound more intuitive with the contrapositive formulation. (By the way, water is not *sufficient* for making coffee. For one thing, you need coffee grains.)

Rule: "if P then Q" = "P only if Q" = "P is sufficient for Q" = "Q is necessary for P"

Examples:

You can be good at logic only if you practice a lot. = (G ⊃ P)

Practicing a lot is necessary for being good at logic. = (G ⊃ P) = (~P ⊃ ~G)

Getting an A is not necessary for passing the course. = ~(P ⊃ A)

Getting an A is sufficient for passing the course. = (A ⊃ P)

§4.2.5 Words Equivalent to "if and only if"

Since "if and only if" has the word "and" in it, it could also come with any of the above equivalent terms for "and" without change of meaning. So "if but only if", "if however only if", etc. all mean the same thing as "if and only if". Another phrase that is often used synonymously with "if and only if" is "just in case". Notice that "just in case" is sometimes used in everyday English to imply that we are being very discreet in that we are considering all probabilities. So, for example, we say things like "Take your umbrella, just in case it rains". This is *not* a case in which "just in case" means "if and only if". So, once again, be careful not to be deceived by the appearances.

As you saw above, "in case" means "if". So "just in case" and "just if" are equivalent, which means "just if" is also another way of saying "if and only if". But be careful, "just if" does *not* mean the same thing as "only if". Here, "just if" must be understood as meaning "exactly if". It is easier to understand why "exactly if" has the same meaning as "if and only if": When two propositions are joined by the biconditional, there is a mutual relationship between them. In one of our examples above, we talked about what it means to say "Sara goes to the party if and only if James goes". We said that if this is true, you'll never see one of them at the party without the other. So Sara going to the party is not one of the conditions for James to go to the party, but *the only* condition. This is the sense in which we can say "Sara goes to the party *exactly if* James goes".

We saw above that "necessary" goes one way while "sufficient" goes the other way. So if we put them together, we get a two-way relationship. Therefore, "P is *necessary and sufficient* for Q" is the same as "P if and only if Q".

Rule: "if and only if" = "if but only if" = "just in case" = "just if" = "exactly if" = "is necessary and sufficient for"

Examples:

The paper will get accepted if but only if it says what the referees like. = (P ≡ R)

Our company will survive just in case we can stay ahead of our competitors. = (S ≡ A)

She approves of things exactly if her mother does not. = (S ≡ ~M)

Being proven guilty is necessary and sufficient for punishment. = (G ≡ P)

We have seen some equivalent phrases for each of our five basic connectives. In the next couple of sub-sections, we will study some important phrases that are not obviously equivalent to any of our connectives, but can be built using two or more connectives.

§4.2.6 Exclusive or

We said above that our disjunction is always an inclusive or. But we also said that we are still capable of representing the exclusive or in our language. It is time to see how. When we use "or" in the exclusive sense in a sentence like "P or Q", we mean to say that either P or Q happens *but not both*. So, the first way of translating the exclusive or is to follow this intuition.

Rule: "P or Q" (exclusive) = "P or Q, and not both P and Q"

Examples:

She will date either Carl or Ricky (exclusive or). = ((C ∨ R) . ~(C . R))

Your mail will be delivered either tonight or tomorrow night (exclusive or). = $((N \vee T) \cdot \sim(N \cdot T))$

You will see other equivalent translations for the exclusive or later in this course.

§4.2.7 "Unless"

How should we translate sentences involving "unless"? Suppose someone says "I will stop talking to her unless she tells the truth". One thing this definitely implies is that "If she does not tell me the truth, I'll stop talking to her". So, roughly, "unless" corresponds to "if not". But sometimes our intuitions tell us that more is involved. So, for instance, "I will continue to defend my view unless I hear a good argument to the contrary" seems to be saying two things: "If I don't hear a good argument to the contrary, I'll continue to defend my view" and "If I hear a good argument to the contrary, I will not continue to defend my view". So, loosely speaking, sometimes "unless" seems to "go both ways". So there is an ambiguity involved. In everyday contexts, we usually understand whether it goes only one way or both ways based on the content and the context. Since our logical language admits of no ambiguities, we have to make a choice (just like we did in the case of inclusive vs. exclusive or). The usual convention is to interpret "unless" as "going only one way". So, we shall translate "unless" as "if not".

"Unless" might also come at the beginning of the sentence. Once again, don't pay attention to the order of the English sentences. Pay attention to which sentence immediately follows the connective.

There is another way to translate "unless" that is equivalent to "if not". In general, "If not P then Q" is equivalent to "Either P or Q" (inclusive). We will prove this (perhaps surprising) fact later on. But in the case of "unless" the equivalence is more intuitive, especially when "unless" is used in threats: "I'll continue the interrogation unless you tell me where he is" can be paraphrased as "Tell me where he is or I'll continue the interrogation".

Rule: "unless" = "if not" = "or"

Examples:

I will stop talking to her unless she tells the truth. = $(\sim T \supset S) = (T \vee S)$

I will continue to defend my view unless I hear a good argument to the contrary. = $(\sim G \supset C) = (G \vee C)$

I don't want to be with him anymore unless he makes up for what he did before. = $(\sim M \supset \sim B) = (M \vee \sim B)$

I can't meet with you tonight unless you don't go to that party. = $(\sim\sim P \supset \sim M) = (\sim P \vee \sim M)$

It is up to you whether you want to translate "unless" as "if not" or as "or". Notice that if you choose to translate "unless" as "if not", there is always an extra negation that gets applied to the sentence after "unless". If the sentence already has a negation (as in the last example), you must put double negations.

§4.2.8 "Neither… nor…"

Consider the sentence "Neither Democrats nor Republicans represent our values". This is just a more elegant way of saying "Democrats don't represent our views and Republicans don't represent our views". So, in general, "Neither P nor Q" is equivalent to "Not P and not Q". The correct translation for this is $(\sim P \cdot \sim Q)$. It is very important that you put a negation in front of each sentence letter. You cannot save ink by "factoring the negation": $\sim(P \cdot Q)$ is an entirely different claim. While $(\sim P \cdot \sim Q)$ says that neither of the two happens, $\sim(P \cdot Q)$ merely rules out the possibility of both happening. Thus, $\sim(P \cdot Q)$ will be true if, say, P alone happens, while $(\sim P \cdot \sim Q)$ will be false under the same circumstances.

There is, however, another way to express the idea of "neither… nor…". As we said above, $(P \vee Q)$ says that at least one of P or Q happens. This is exactly what "Neither P nor Q" denies. So we can get an equally good translation of "Neither P nor Q" by negating $(P \vee Q)$.

Rule: "Neither P nor Q" = "Not P and not Q" = "It is not the case that either P or Q"

Examples:

I'll eat neither the spaghetti nor the pizza. = $(\sim S \cdot \sim P) = \sim(S \vee P)$

I am neither a theist nor an atheist. = $(\sim T \cdot \sim A) = \sim(T \vee A)$

§4.2.9 Complex Sentences

Sentences of English sometimes involve several connectives. You might, for example, have "not", "if… then…" and "or" in one sentence. By "complicated" we don't mean hard to understand or lengthy here, but involving multiple connectives. So the following sentence is logically very simple:

> The essential properties of an object of synthetic *a priori* knowledge are to be sure partly dependent for their revelation on the contingencies of an enterprise that is in and of itself empirical.

The above sentence is logically very simple because, from the point of view of PL, it is just one atomic sentence, so we can translate it as, say, P. It has no structure, i.e. no negation, conjunction, disjunction, conditional, or biconditional, in it (and you don't have to *understand* the sentence to realize this!). But the following sentence is logically quite complicated:

> Assuming I love him and he doesn't love me back, we will either not be happy together or if we will, we won't stay with each other forever.

The complexity lies in the existence of several connectives in one sentence. To be able to translate such sentences, you need to understand the notion of the *main connective*. It is hard to define what the main connective is but it is usually easy to find it intuitively. So, for instance, consider this sentence: "If Eric is handsome and generous, I'll be interested in him". There are two connectives in this sentence, namely "if… then…" and "and". But the main connective is obviously the conditional. The statement is telling us the conditions under which the person will be interested in Eric, while the conditions happen to be twofold. The comma usually helps locate the main connective in complex sentences. Since there might be several commas in one sentence, you might need to read the sentence several times, pausing at various places where commas go. The most natural-sounding pause is most likely where the main connective is. In any case, the main connective is the one that connects the largest "chunks" to each other. If you were building the sentence from the smallest units, the main connective would be the one that you'd apply last.

The main connective is the connective that embraces the largest chunks of a sentence. To find it, read the sentence with pauses at various places. The pause that sounds most natural is where the main connective goes.

Here are some more examples:

Sentence: "With the dish comes local bread and either a local drink or a coke". Main connective: "and". Translation: (B **.** (D ∨ C))

Sentence: "I am not stingy and mean". Main connective: "not". Translation: ~(S **.** M)

Sentence: "Although I didn't know him that well, I felt comfortable and happy in his company". Main connective: "although" (= "and"). Translation: (~K **.** (C **.** H))

The following is a step-by-step strategy for tackling complex sentences:

1- Underline all the connectives.

2- Identify the main connective.

3- Paraphrase the main connective if necessary.

4- Write down the appropriate logical symbol for the main connective with "slots" for the chunks that it connects.

5- Treat the "chunks" that the main connective is applied to as new sentences. If they are simple (atomic) sentences, assign sentence letters to them. If they involve further structure, translate them by repeating 1-5.

Let us apply the above strategy to our example of a complex sentence above:

Assuming I love him and he doesn't love me back, we will either not be happy together or if we will, we won't stay with each other forever.

First, we underline all the connectives.

> Assuming I love him and he doesn’t love me back, we will either not be happy together or if we will, we won’t stay with each other forever.

Second, we identify the main connective. The sentence mainly tells us what happens if the speaker loves some person and he doesn’t love him/her back. So the main connective is “assuming”. Another way to see this is by locating the commas and realizing that “assuming” connects the largest phrases together. You can also apply the “pause method”.

In any case, “assuming” must be paraphrased as “if… then…”. Now that we know the main connective and have appropriately paraphrased it, we write it down, with slots for the chunks that go before and after.

([something] ⊃ [something])

The chunk that goes before the horseshoe is “I love him and he doesn’t love me back”. The main connective of this sentence is the “and”, so we can translate it as (I • ~B). Thus, so far we have:

((I • ~B) ⊃ [something])

The chunk that goes after the horseshoe is slightly more complicated: “We will not be happy together or if we will, we won’t stay with each other forever”. Let us apply our method anew to this sentence. The main connective is “or”. The reason is that it doesn’t make much sense to pause at the comma: “We will not be happy together or if we will [pause…] we won’t stay with each other forever”. The part that comes before the pause cannot be a complete sentence on its own. On the other hand, it makes perfect sense to pause before “or”: “We will not be happy together [pause…] or if we will, we won’t stay with each other forever”. So, for this sentence we have:

([something] ∨ [something])

Before the "or" we have "We will not be happy together", which can be translated as ~H. The chunk after the "or" is "If we will [be happy together], we won't stay with each other forever". The translation for this would be (H ⊃ ~S). Therefore, we have:

$(\sim H \vee (H \supset \sim S))$

Plugging this back in the original sentence, we finally get:

$((I \cdot \sim B) \supset (\sim H \vee (H \supset \sim S)))$

The best way to know you translation is correct is to read your symbolic sentence for yourself, replacing symbols with actual sentences, and see if the result has the same content as the original English sentence or not.

Having learned how to translate English sentences into PL, we are ready to study some interesting methods for analyzing statements and arguments logically in the following chapters.

§4.E Homework

HW6

1- (6 pts.) Complete CET, CHM, and CHT from LogiCola.

2- (0.75 pt. each) Translate the following sentences into PL.

a. Neither Tim nor Kevin will attend the meeting, unless Bradley accepts their conditions.

b. I'm not going to let him get away so easily, given that he is so rude and doesn't apologize or make up for what he did.

c. Although they talk and laugh and help each other if needed, assuming what I heard is correct, they're not in good terms,.

d. I'm sorry I can't help you, but if you want, you can talk to my boss and tell him what happened.

3- (0.5 pt. each) Decide whether inclusive or exclusive or is meant and translate the sentence accordingly.

a. You can either email or fax the documents.

b. Tonight, either we'll win the battle or we'll all meet each other in heaven.

Further Reading:

Gensler, Harry J. (2002), *Introduction to Logic*, Routledge, Chapter 3, Sections 3.1, 3.2, and 3.8

Teller, Paul (1989), *A Modern Formal Logic Primer*, available online, Volume I, Chapters 1 and 2

Merrie Bergmannn, James Moor, Jack Nelson (2009), *The Logic Book*, the McGraw-Hill Companies, Chapter 2, Sections 2.1 and 2.2

Chapter 5

Propositional Logic: Truth Tables

§5.1 How to Construct Truth Tables

We have already seen five basic truth tables for our five basic connectives in the previous chapter. In this chapter, we will build truth tables for sentences with multiple connectives and use them for a variety of purposes. We shall see that truth tables are very useful tools for determining the logical properties of statements, pairs or sets of statements, and arguments.

To construct a truth table for a sentence with multiple connectives, we take our five basic truth tables as given, and use them in a step by step process to fill out the truth table. The goal of each exercise is to find the *truth values under the main connective*. So, let us start with a simple example. Consider the sentence (P ⊃ ~P). The main connective of the sentence is the conditional. First, we fill out the left-hand side of the truth table (which is always the same), remembering that for one sentence there are only two rows:

P	(P ⊃ ~P)
0	
1	

Now, to find the truth values under the horseshoe, we need to have the truth values on the two sides of the horseshoe. So, we fill them out, using our knowledge of the truth table for negation:

P	(P	⊃	~P)
0	0		1
1	1		0

Since the antecedent is just P, we simply copied the values from the left-hand side. But for the consequent, ~P, we applied negation. And, finally, we fill out the values under the conditional, using the basic truth table for the conditional connective.

P	(P ⊃ ~P)
0	0 **1** 1
1	1 **0** 0

The **bold** font is meant to indicate that this is our final result, the values under the main connective. Here is how we found the answer: in the first row we have (0 ⊃ 1), which, according to the truth table that defines horseshoe (see Chapter 4), gives 1. In the second row, we have (1 ⊃ 0), which, according to the same truth table, gives 0.

Let's try another example, this time with two sentences in it. Consider the following sentence: ((P ∨ Q) **.** ~(P **.** Q)). You remember from Chapter 4 that this is the "exclusive or". It says that either P or Q can happen but rules out the possibility of both happening. Since we have two sentences, we will have four rows in our truth table as follows.

P	Q	((P ∨ Q) . ~(P . Q))
0	0	
0	1	
1	0	
1	1	

The main connective is the dot in the middle (the first "and"). We know the basic truth tables for disjunction and conjunction, so we fill them out:

P	Q	((P ∨ Q) . ~(P . Q))	
0	0	0	0
0	1	1	0
1	0	1	0
1	1	1	1

Then we apply the negation to the result of the last column (see next page). This reverses all the truth values on the last column. From this point on we can forget the values on the last column, because we will be working only with the values under the negation.

P	Q	((P ∨ Q) . ~(P . Q))
0	0	0 1 0
0	1	1 1 0
1	0	1 1 0
1	1	1 0 1

And, lastly, we apply the conjunction to the two columns on its sides, i.e. to the values under the disjunction and the negation. We have:

P	Q	((P ∨ Q) . ~(P . Q))
0	0	0 **0** 1 0
0	1	1 **1** 1 0
1	0	1 **1** 1 0
1	1	1 **0** 0 1

We found the answer by noting that (0 . 1) gives 0, (1 . 1) gives 1, and (1 . 0) gives 0. The result makes perfect sense in terms of what we expect from the exclusive or. According to the bold column, the sentence ((P ∨ Q) . ~(P . Q)) is only true when exactly one of P or Q happens. It is false when neither or both happen. This is just what a statement with exclusive or says.

Here is another example of a truth table with two atomic sentences.

C	G	~((C ≡ ~G) ∨ C)
0	0	**1** 0 0 1 0 0
0	1	**0** 0 1 0 1 0
1	0	**0** 1 1 1 1 1
1	1	**0** 1 0 0 1 1

Make sure you can follow the value assignments. Notice that the main connective does not have to be "in the middle". It can be the first column, if it's a negation, as it is in the above case.

As our last example, let us consider a sentence with three atomic sentences: ((A . B) ∨ ~C). For three sentences, we need eight rows, because there are 2 × 2 × 2 possibilities. These possibilities can be arranged in a systematic way in the following manner.

A	B	C	
0	0	0	
0	0	1	
0	1	0	
0	1	1	
1	0	0	
1	0	1	
1	1	0	
1	1	1	

The first sentence gets four 0's and four 1's, because four is half of eight. Then the second sentence gets two 0's and two 1's alternately, and the third sentence one 0 and one 1 alternately. The rest of the work is not any more difficult (but slightly more time-consuming) than when we had two sentences (i.e. four rows). All we need is patience and carefulness. Here is the result:

A	B	C	((A . B)	∨	~C)
0	0	0	0	**1**	1
0	0	1	0	**0**	0
0	1	0	0	**1**	1
0	1	1	0	**0**	0
1	0	0	0	**1**	1
1	0	1	0	**0**	0
1	1	0	1	**1**	1
1	1	1	1	**1**	0

To fill out the values under the conjunction, we used the fact that (A . B) is only true if both A and B are true. This only happens in the last two rows. (A . B) is false everywhere else.

So far, we learned how to construct truth tables. In the following few sections, we will learn some of the great uses of truth tables.

§5.2 Logical Truth, Logical Falsity, and Logical Indeterminacy

Statements that tell us something about the world can be either true or false. So "Santa Claus exists", "Water boils at a 100 degrees Celsius" and "If you're 5 years old, then you can't have children" are all statements that can be true or false. Some of them might be true in our actual world, while others might be false, but they are all such that we can *imagine* a world in which they are true and a world in which they are false. When we construct truth tables, that is basically what we are doing. Each row of the truth table corresponds to a (hypothetical) world where our statement is either true or false. Even if water never boils at other temperatures throughout the history of the world, it is still possible to imagine that it does.

But there are statements that can never be false. "If you're 5 years old, then you're 5 years old" is one such statement. You cannot even imagine a hypothetical world in which this is false. The reason is that this statement is not true because of how things turned out to be in our world, say how the human physiology happens to work and so on, but by virtue of its logical structure. Any statement of the form "If P, then P" has the same property of being necessarily true, in the actual world and in all other possible worlds. We call such a statement a *logically true statement*. Make sure you know the difference between being actually true (e.g. "Water boils at 100 degrees Celsius"), and being logically true (e.g. "If P then P").

A statement is logically true if and only if there is no possibility that it is false.

Similarly, we can think of statements that cannot possibly be true, such as "The door is open and not open at the same time". Such a statement expresses a contradiction which can never be true.

A statement is logically false if and only if there is no possibility that it is true.

Most statements that we deal with in everyday life and in the sciences are neither logically true nor logically false. A statement that is neither logically true nor logically false is called a *logically indeterminate* statement. The above examples of "Santa Claus exists", "Water boils at 100 degrees Celsius", and "If you're 5 years old, then you can't have children" are all logically indeterminate statements.

A statement is logically indeterminate if and only if it is neither logically true nor logically false.

The definitions of logical truth and logical falsity employ the word "possibility". This needs interpretation. I said above that you can think of these possibilities as possible worlds that are available to imagination. But in our system of PL, possibility has a much simpler interpretation. A possibility corresponds to a row in the truth table of the statement.

In PL, a "possibility" means a row in the truth table.

So, our definitions of logical truth, falsity, and indeterminacy have the following PL-interpretations.

A statement is logically true in PL if and only if all of its truth values (under the main connective) are 1.

A statement is logically false in PL if and only if all of its truth values (under the main connective) are 0.

A statement is logically indeterminate in PL if and only if its truth values (under the main connective) are a mix of 1's and 0's.

Let us see some examples.

Consider this sentence: (P **.** ~P). This can be the translation of our example above, "The door is open and not open at the same time". Here is the truth table for this sentence:

P	(P **.** ~P)
0	0 **0** 1
1	1 **0** 0

Our intuitions are confirmed. All the values under the main connective of (P **.** ~P) are 0's, meaning that there is no possibility (no row in the truth table) in which it is true.

What about (P ∨ ~P)? One statement that has this structure would be "The door is either open or not open". Well, that sounds like an obvious truth. It can either turn out this way or that way. So we don't need to investigate the laws of nature to determine that this statement is true. Thus it must be logically true. The following truth table confirms our intuitions.

P	(P ∨ ~P)
0	0 **1** 1
1	1 **1** 0

All values under the disjunction are 1's, showing that the statement is logically true.

The above examples were very simple. Don't be misled by them into thinking that being logically true is the same as being *obviously true*. Not all logical truths are obvious. The most complicated mathematical theorems consist of logically true statements. But you cannot tell that their statements are true just by looking at them. You have to do a lot of calculation. Nevertheless, we want to say that even those complicated logical truths don't say anything about how the world is. They are true in every possible world. Here is an example of a logically true statement that is not so obvious: (((P ⊃ Q) **.** (~Q ∨ P)) ⊃ (~P ≡ ~Q)). It is left as an exercise for you to show that this statement is logically true.

Let's construct the truth table of the following statement: ((P ⊃ Q) ∨ (Q ⊃ P)). In English, the sentence says "Either if P then Q or if Q then P", which means either of P or Q is conditioned upon the other. For example, the sentence "Either if the sky is blue, Mars has three moons, or if Mars has three moons, the sky is blue" has this structure. This does not sound like a logically true statement. It seems that the color of the sky and the moons of Mars are completely unrelated things. Let us see what the truth table says:

P Q	((P ⊃ Q) ∨ (Q ⊃ P))
0 0	1 **1** 1
0 1	1 **1** 0
1 0	0 **1** 1
1 1	1 **1** 1

The statement is logically true! So this is another example of a not-so-obvious logical truth. In fact, this is not only not obvious, but extremely surprising. This result has been so surprising to some logicians that they called the above truth table *the paradox of the material conditional* (the "material conditional" being another name for our horseshoe). Before moving on, let me explain why this "paradox" occurs. The above result can be traced back to our definition of the horseshoe as a truth-functional connective, i.e. a connective that can have a truth table. We said then that this forces us to give up some of the ordinary connotations of "if… then…". In particular, we said that our conditional does *not* represent the relation of cause and effect. The reason the above truth table seems surprising, however, is precisely that we tend to read causality into if-then statements. When we hear "If the sky is blue, Mars has three moons", we are prone to thinking that according to the statement, the color of the sky *causes* the number of the moons of Mars. But when this "if" is simply the horseshoe, that is *not* what the statement says. So the truth table above is not really a paradox, but just another demonstration that our conditional connective does not represent cause and effect.

Let's wrap up by an example of logical indeterminacy. Suppose you're given the following statement: (((G ⊃ J) **.** (J ⊃ L)) ≡ ~J). You might not know if this is logically true or false or indeterminate. The truth table helps us find out.

G	J	L	(((G ⊃ J) . (J ⊃ L)) ≡ ~J)				
0	0	0	1	1	1	**1**	1
0	0	1	1	1	1	**1**	1
0	1	0	1	0	0	**1**	0
0	1	1	1	1	1	**0**	0
1	0	0	0	0	1	**0**	1
1	0	1	0	0	1	**0**	1
1	1	0	1	0	0	**1**	0
1	1	1	1	1	1	**0**	0

The statement is therefore logically indeterminate, because it has both 1's and 0's under its main connective.

§5.3 Logical Equivalence

We saw above that both (~P **.** ~Q) and ~(P ∨ Q) are equally good translations of "Neither P nor Q". We used intuitive arguments to show that the two sentences mean the same thing. But if they mean the same thing, they must always have the same truth value, i.e. it can never happen that one is true and the other is false. After all, they have one and the same content. So logicians say that these two statements are *logically equivalent*.

> *Two given statements are logically equivalent if and only if there is no possibility that one is true and the other is false (they are either true together or false together).*

Note that logical equivalence is a property of *pairs of statements*. It cannot be attributed to a single statement, to an argument, or to anything else.

Just as we did in the cases of logical truth, logical falsity, and logical indeterminacy, we can give a PL-adapted interpretation of the above definition.

> *Two given statements are logically equivalent in PL if and only if all of their truth values (under their main connectives) in their joint truth table are the same.*

A joint truth table is a truth table in which we put the two statements side by side. Like this:

P Q	(~P . ~Q)	~(P ∨ Q)
0 0	1 **1** 1	**1** 0
0 1	1 **0** 0	**0** 1
1 0	0 **0** 1	**0** 1
1 1	0 **0** 0	**0** 1

We compare the truth values under the main connectives (the bold columns) to determine if the two sentences are logically equivalent or not. The rest of the 0's and 1's don't matter. The truth table above proves that (~P . ~Q) and ~(P ∨ Q) are logically equivalent, thereby confirming our intuitive arguments about "neither … nor …" statements in the previous chapter.

Let's work out an example with 3 sentence letters. Are ((F . G) ∨ H) and (F . (G ∨ H)) logically equivalent? The two sentences look the same except the parentheses are in different places.

F	G	H	((F . G)	∨	H)	(F .	(G ∨ H))	
0	0	0	0	**0**	0	0 **0**	0	
0	0	1	0	**1**	1	0 **0**	1	←
0	1	0	0	**0**	0	0 **0**	1	
0	1	1	0	**1**	1	0 **0**	1	←
1	0	0	0	**0**	0	1 **0**	0	
1	0	1	0	**1**	1	1 **1**	1	
1	1	0	1	**1**	0	1 **1**	1	
1	1	1	1	**1**	1	1 **1**	1	

So the two sentences are *not* logically equivalent, because of rows 2 and 4. In general, do not expect two sentences to be equivalent if the only difference between them is where the parentheses go. Parentheses are very important for determining what the sentence is saying and shifting them around can change the meaning of the statement entirely. In the above case, the first sentence is saying that either both F and G happen or else H happens. The second sentence, however, says that F definitely happens and either of G or H happens. So, for one thing, the second sentence guarantees the occurrence of F while the first sentence does not.

Here are some very important logical equivalence rules that you need to learn and be comfortable with.

De Morgan's laws:

~(P ∨ Q) = (~P . ~Q)

~(P . Q) = (~P ∨ ~Q)

The Law of Contraposition:

(P ⊃ Q) = (~Q ⊃ ~P)

The relation between the conditional and disjunction:

(P ⊃ Q) = (~P ∨ Q)

The relation between the biconditional and the conditional:

(P ≡ Q) = ((P ⊃ Q) . (Q ⊃ P))

Be careful, the "=" sign is *not* a connective. I am simply using it as shorthand for "is logically equivalent to".

§5.4 Truth Table Test for Validity

Another very useful application of truth tables is in determining whether a deductive argument is valid or not. We have already seen several examples of deductive arguments and have used our intuitions to decide about their validity. But truth tables provide indisputable proofs for validity or invalidity of any given argument that can be translated into PL.

Remember the definition of validity.

An argument is valid if and only if there is no possibility that all premises are true and the conclusion is false.

Since "possibility" in PL means a row in the truth table, we will have the following PL-adapted definition for validity.

An argument is valid in PL if and only if there is no row in the truth table in which all premises are true and the conclusion is false.

Thus, to determine whether an argument is valid or not, we construct a joint truth table in which we put all the premises and the conclusion side by side. We are looking for a line that makes the argument invalid. If we find it, we have proven the argument invalid. If we don't find it, it is valid.

Consider the argument on the next page. This is the quintessential valid argument; if any argument is valid, this one is! It has been known since antiquity to be a deductively valid form of reasoning and so has its own Latin name: *modus ponens*. It says that when P is a sufficient condition for Q, and the condition obtains (P is true), the consequent (i.e. Q) cannot but be true. Anyway, our goal is to see if the truth table confirms that *modus ponens* is valid. The truth table is shown on next page. We use the sign $\therefore$ to designate the conclusion.

$(P \supset Q)$

P

Q

P Q	(P ⊃ Q)	P	∴ Q
0 0	**1**	**0**	**0**
0 1	**1**	**0**	**1**
1 0	**0**	**1**	**0**
1 1	**1**	**1**	**1**

Notice how we determine validity. We are looking for a row with 1 1 0 as the values under the main connectives of the three sentences. If we find 1 1 0, the argument is *invalid*. If we don't find it, it is valid. We *do not* look for a row that proves the argument valid. There will be no such row. You cannot prove that the argument is valid directly. You can only prove validity by showing that invalidity (i.e. 1 1 0) does not occur.

The following is a particularly common mistake. When students find a row that has 1 1 1 (all true) under the main connective, they often think that the argument is automatically valid. Make sure you don't make the same mistake. Remember from Chapter 1 that validity is *not* about truth or falsity of either the premises or the conclusion. A valid argument may or may not have true premises and conclusion. What determines validity is the *im*possibility of *the conclusion being false when the premises are true*, which, in terms of our system of PL, amounts to there being no row where 1 1 0 occurs.

When the argument has three premises, we will be looking for 1 1 1 0. If it has four premises, we are looking for 1 1 1 1 0. And so on. The most common arguments have two premises, but there is no restriction on the number of premises of an argument.

Strategy: Look for 1 1 … 1 0 (all 1's and a 0 under the main connective of the conclusion). If you find it, the argument is invalid*. If not, it is valid. Don't go after a row that proves the argument to be valid.*

Don't confuse the number of premises with the number of atomic sentence letters. The number of premises determines the number of partitions in your truth table. But the number of atomic sentences determines the number of rows. The argument below has three premises but two atomic sentences:

(P ≡ Q)

(Q ≡ ~P)

P

~P

The following argument, however, has two premises but three sentence letters:

(~F ∨ G)

(G ⊃ ~H)

(~F ≡ H)

Let's examine the validity of both of the above examples. For the first argument we have:

P Q	(P ≡ Q)	(Q ≡ ~P)	P	∴ ~P
0 0	**1**	0 **0** 1	**0**	**1**
0 1	**0**	1 **1** 1	**0**	**1**
1 0	**0**	0 **1** 0	**1**	**0**
1 1	**1**	1 **0** 0	**1**	**0**

We are looking for 1 1 1 0 which cannot be found in any row in the above truth table. Therefore, the argument is valid. When you get used to this type of exercise, you can save a lot of time by skipping the rows where the pattern in question cannot possibly occur. So, for example, in the second row above, we have a 0 under the main connective of the first premise. Therefore, 1 1 1 0 cannot possibly happen in this line. Since we know this, we will not have to fill out the rest of this row. But it is less confusing at the beginning to fill out the truth table thoroughly.

Let's examine the second argument given above.

F G H	(~F ∨ G)	(G ⊃ ~H)	∴ (~F ≡ H)	
0 0 0	1 **1** 0	0 **1** 1	1 **0** 0	←
0 0 1	1 **1** 0	0 **1** 0	1 **1** 1	
0 1 0	1 **1** 1	1 **1** 1	1 **0** 0	←
0 1 1	1 **1** 1	1 **0** 0	1 **1** 1	
1 0 0	0 **0** 0	0 **1** 1	0 **1** 0	
1 0 1	0 **0** 0	0 **1** 0	0 **0** 1	
1 1 0	0 **1** 1	1 **1** 1	0 **1** 0	
1 1 1	0 **1** 1	1 **0** 0	0 **0** 1	

The argument is thus invalid, because of the rows indicated by the arrow.

We can now combine the skills that we learned in Chapter 4 and this chapter to translate and analyze real-life arguments. For example, consider this argument:

> In our society, if you are proven guilty you will be punished. But punishment is a cruel act that is sufficient for the subject no longer having a human status. Thus, in our society, if you are not guilty, you are a human, but if you are guilty, you are not a human.

We recognize the *conclusion* by such words as "thus", "therefore", and the like. So the last sentence is the conclusion. The first premise can be summarized as "If proven guilty then punished", which could be translated as (F ⊃ G). The second premise of the argument can be summarized as "If punished then not human", which would be (G ⊃ ~H). The conclusion says "If not guilty then human and if guilty then not human". This can be translated as ((~F ⊃ H) **.** (F ⊃ ~H)). These translations are put in the form of an argument on the next page. It is left as an exercise to show that this argument is invalid. But if you're good at equivalence rules, you can see that this is the exact same argument as the one we just proved to be invalid. Here is why. By one of the equivalence rules that we learned above, the first premise is the same as (~F ∨ G). The second premise is identical to what we had in the example above. As for the conclusion, by the Law of Contraposition, (F ⊃ ~H) is equivalent to (H ⊃ ~F); so the conclusion can be written as ((~F ⊃ H) **.** (H ⊃ ~F)). But we know that this is equivalent to (~F ≡ H).

(F ⊃ G)

(G ⊃ ~H)

((~F ⊃ H) . (F ⊃ ~H))

Of course this argument was designed to turn out the same as our previous example; but in general, knowing the equivalence rules (especially the important ones) is greatly helpful in recognizing identical patterns of argument and converting a given argument to one that we know to be valid/invalid, instead of evaluating each and every argument separately. For instance, once we know that an argument with (P ⊃ Q) as the premise and (Q ⊃ P) as the conclusion is invalid, we automatically know that an argument with the same premise but with (~P ⊃ ~Q) as the conclusion is also invalid, because the latter is the contrapositive of (Q ⊃ P).

§5.E Homework

HW7

1- (4 pts.) Complete DFM and DFH from LogiCola.

2- (0.5 pt. each) Determine whether each of the following statements is logically true, logically false, or logically indeterminate.

a. ~(A ⊃ (B ⊃ A))

b. ~((A ⊃ B) ⊃ A)

c. (((P ⊃ Q) . (~Q ∨ P)) ⊃ (~P ≡ ~Q))

d. (((W ∨ V) . (W ∨ Y)) ⊃ ~(V ∨ Y))

3- (0.25 pt. each) Use truth tables to prove all equivalence rules given at the end of Section 5.3, except for De Morgan's *first* law which was proved in the text (include the second law).

4- (0.5 pt.) Explain De Morgan's second law in intuitive terms (use examples and the right wording to show that the two sentences have the same message).

5- (0.5 pt.) Explain the Law of Contraposition in intuitive terms (use examples and the right wording to show that the two sentences have the same message).

6- (0.5 pt. each) Use a truth table to determine whether the following pairs are logically equivalent.

a. (P ⊃ (Q ⊃ R)) and ((P . Q) ⊃ R)

b. ((~A ∨ B) . C) and (~A ∨ (B . C))

c. ((F . ~(G ∨ C)) ⊃ ~G) and ((G ∨ ~(F . C)) ⊃ F)

d. (~A ⊃ ~A) and (C ≡ C)

HW8

1- (6 pts.) Complete DAM and DAH from LogiCola.

2- (0.75 pt.) Show that the argument regarding the effects of punishment given at the end of Section 5.4 is invalid with the translation given on p. 79.

3- (0.5 pt.) Show that the same argument would be valid if the conclusion were just "In our society, if you are guilty, you are not a human".

4- Consider *modus tollens*, another famous form of argument which reads: (P ⊃ Q), ~Q, ∴ ~P.

a. (0.25 pt.) Use a truth table to show that *modus tollens* is valid.

b. (0.5 pt.) Explain why *modus tollens* is valid in intuitive terms (how would you convince someone that not P, if they already accepted that if P then Q and that not Q?).

5- (2 pts.) Translate this argument into PL and determine whether it is valid or invalid.

If you subscribe to the values of the political left, you deeply care about the poor. But not embracing these values is a necessary condition for being a Republican. Therefore, you can be a Republican only if you don't care about the poor.

Further Reading:

Gensler, Harry J. (2002), *Introduction to Logic*, Routledge, Chapter 3, Sections 3.5 and 3.6

Teller, Paul (1989), *A Modern Formal Logic Primer*, available online, Volume I, Chapters 3 and 4

Merrie Bergmannn, James Moor, Jack Nelson (2009), *The Logic Book*, the McGraw-Hill Companies, Chapter 3

Chapter 6

Propositional Logic: Proofs

§6.1 The Idea of A Proof System

In Chapter 5 we learned how to decide whether an argument is valid or invalid. Thus our system of logic can be considered a model of logical reasoning insofar as it can decide between valid and invalid arguments and insofar as its decisions reflect those of rational people fairly well. But when we make arguments in real life, usually we don't just provide some premises and then jump to the conclusion and claim the move to be valid. Rather, when the argument is sophisticated, we go through a few steps that take us from the pieces of information given to us in the premises to the final conclusion. We might therefore like to have a system of deduction that represents this step-by-step process of thinking through an argument.

As an example, consider *modus tollens* again, a form of argument that we showed to be valid in an exercise in Chapter 5.

$(P \supset Q)$

$\sim Q$

$\sim P$

We know it is valid, but how do we get from the premises to the conclusion? When one is reasoning through a *modus tollens*, it usually goes something like this: "Well, I know that Q didn't happen. But I also know that if P had happened, Q would definitely have happened. So, in retrospect P must have failed to happen as well, because otherwise we would end up with Q, which we didn't". It seems that what is going on here is along the following lines:

1- Assume for a moment that P happened.

2- From premise 1 we know that Q would then have to happen.

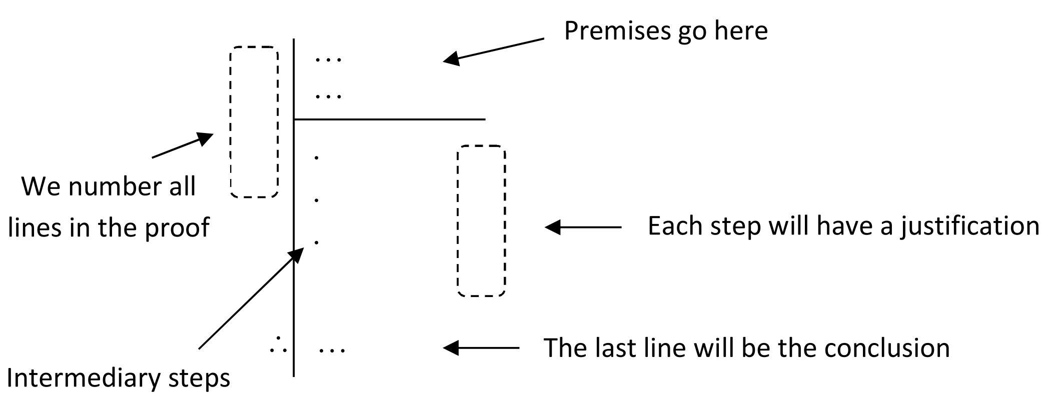
3- But from premise 2 we know that Q didn't happen.

4- Since it cannot be that Q both did and did not happen, our tentative assumption that P happened must be false.

We would like to have a logical system that not only tells us *modus tollens* is valid, but also mimics the step-by-step reasoning shown in 1-4 above. Such a system is called a *natural deduction* system. A natural deduction system consists of a small number of simple (and preferably intuitive) rules that we take for granted, that allow us to prove all other, more sophisticated arguments. There are several systems of natural deduction out there, each of which has certain advantages and disadvantages. We will study one of them here.

The nice thing about the natural deduction system we will study below is that its rules are directly based on our connectives. In this system, we will have two rules per connective, an *elimination rule* and an *introduction rule*. These add up to 10 rules. We will also have one simple rule called *Reiteration*. Thus, our natural deduction system has 11 basic rules. We will be able to prove *all* valid arguments with these 11 rules. To save some time, however, we shall skip Biconditional Elimination and Biconditional Introduction in this course. If you ever want to prove an argument that has biconditionals in its premises or conclusion, use the equivalence rules to convert them into simple conditional claims and then use the conditional rules explained below.

An argument with all of its steps explicitly shown is called a *proof*. We *derive* the conclusion from the premises. For this reason, proofs are also called *derivations.* We will call them with both names. To show an argument in the previous chapters, we always simply listed the premises and separated the conclusion from them by a horizontal line. This notation does not leave any space for working out our step-by-step calculations. So, we will adopt the following notation from now on. While a horizontal line still separates the premises from everything else, we also draw a *vertical* line that runs through our reasoning.



Here is a sample proof (you are not expected to understand what is going on in this proof, but just to get your eyes familiarized with where everything goes):

1.	(A • B)	A
2.	A	1, •E
3.	(A ∨ C)	2, ∨I
4.	B	1, •E
5.	(D ∨ B)	4, ∨I
∴	((A ∨ C) • (D ∨ B))	3, 5, •I

These are justifications

This is the conclusion

The *justification* in front of each sentence is a short, notational explanation of where from and how we got the given sentence. The premises don't need justification, because they are our given assumptions. So we just put an uppercase A in front of them, to stand for "assumption". The conclusion is designated by the triangular symbol from the last chapter. We are now ready to learn our basic rules. Let us start with Reiteration.

§6.2 Reiteration

Reiteration is a very simple rule. It simply says that we are allowed to repeat a sentence that we have already proven in our derivation, if we need to. This is allowed because it is always valid to conclude P from the assumption of P itself. We can demonstrate this rule as follows.

	⋮	
i.	P	
	⋮	
∴	P	i, R

The Rule of Reiteration

Here is how to understand the above demonstration. If you have proven or assumed P somewhere in your proof, you can always write down P later on in your proof.[1] We cite the original line and the Reiteration rule to justify it. As you see, the justification reads "i, R", which means we used line i and the Reiteration rule. Of course, in an actual example, a line number (such as 3, 7, etc.) will take the place of i. The dots don't mean anything. They just show that there can be other miscellaneous lines in between. You might wonder what Reiteration is good for. We will see its applications later on.

One thing you need to know about Reiteration is that it only allows you to copy and paste the sentence *exactly as it is*. No change in the symbols is allowed, even if it does not change the meaning of the sentence. For example, you cannot make either of the following moves:

1.	(A . B)	A
∴	A	1, R

Incorrect application of R

1.	(A . B)	A
∴	(B . A)	1, R

Incorrect application of R

However obvious it seems to you that A can be derived from (A . B) or that (B . A) is the same as (A . B), we will not take those for granted until we prove them.

§6.3 Conjunction Elimination

Consider this argument:

Both Sara and James went to the party.

Therefore, Sara went to the party.

[1] There is a caveat to this that you'll see soon.

This is certainly valid. Any argument that has (P . Q) as its premise and P as its conclusion is valid. Apart from the fact that a truth table can show that this argument is valid, its validity is intuitively obvious. Conjunction Elimination is a rule that allows you to make this move in your arguments, if you need to. The reason we call it Conjunction Elimination is that the premise contains conjunction but the conclusion typically does not. So it is as if we "eliminated" the conjunction. The rule is demonstrated below.

```
   |
   |________
   | .
   | .
   | .
i. | (P . Q)
   | .
   | .
   | .
∴  | P            i, .E
```

The Rule of Conjunction Elimination

```
   |
   |________
   | .
   | .
   | .
i. | (P . Q)
   | .
   | .
   | .
∴  | Q            i, .E
```

The Rule of Conjunction Elimination

The rule states that if you have proven or assumed (P . Q) somewhere in your proof, you can write down either of P or Q later on. We write .E in the justification and read it as "conjunction-elimination" or, more casually, "and-elimination" or "dot-elimination". Again, i is simply a place-holder for an actual line number and the horizontal dots merely show that there could be other lines in between. The rule allows the derivation of P and of Q from (P . Q). Which of P or Q you should derive depends on what you eventually want to do in the proof.

So far we have learned two rules, namely Reiteration and Conjunction Elimination. There aren't many interesting proofs that we can write down using only these two rules. But if you are eager to see an actual proof that you can fully understand, there is one below. This is of course a quite boring proof and the repeating of A in line 3 seems pointless (if your desired conclusion is A, you can just stop at line 2, there is no reason to repeat it). But this proof was just for the purpose of demonstration. We will see many interesting and at times difficult proofs later on.

1.	(A . B)	A
2.	A	1, .E
∴	A	2, R

§6.4 Conjunction Introduction

As said above, each connective has two rules associated with it, an elimination rule and an introduction rule. We learned Conjunction Elimination. So let's see what Conjunction Introduction looks like. Consider this argument:

Pigeons are birds.

Ravens are brids.

Therefore, Pigeons and Ravens are birds.

It is surely a valid argument. The conclusion adds together the pieces of information given in the premises. Conjunction Introduction is the rule that allows you to do so. It is called Conjunction Introduction because the conclusion contains conjunction while the premises need not. It is therefore as if we "introduced" a conjunction into our sentence.

As shown on the next page, the rule allows us to derive (P . Q) from two separate statements, P and Q. Deriving (Q . P) is also legalized by the same rule. Once again, which one you want to derive depends on your goals in that particular derivation. When using Conjunction Introduction, you have to cite *two* line numbers. The order of the numbers does not matter, but you have to cite both lines. This is very important for the correctness of your proof. If you only cite line j, for example, you are practically claiming that you derived (Q . P) from Q alone, which is obviously invalid.

	⋮	
i.	P	
	⋮	
j.	Q	
	⋮	
∴	(P . Q)	i, j, .I

The Rule of Conjunction Introduction

	⋮	
i.	P	
	⋮	
j.	Q	
	⋮	
∴	(Q . P)	i, j, .I

The Rule of Conjunction Introduction

With the two conjunction rules that we have learned so far, we can carry out the following proof:

1.	(R . S)	A
2.	R	1, .E
3.	S	1, .E
∴	(S . R)	2, 3, .I

The goal of the argument is to derive (S . R) from (R . S). We basically broke it apart and put it back together. You might wonder why it would be necessary to prove such a thing at all. Isn't (S

. R) the same as (R . S)? Indeed, the two sentences have the same content; they both claim that both S and R happened. They are, however, two different strings of symbols. If, for example, one were you password, you couldn't log in by typing the other one. And, as was said above, in our logic we don't assume that two strings of symbols are the same unless we can prove that.

§6.5 Conditional Elimination

Let's turn to the conditional rules. The rule of Conditional Elimination corresponds to the following argument.

If Smith accepted bribes, then he is corrupted.

Smith did accept bribes.

Therefore, Smith is corrupted.

The rule states that if you know that (P ⊃ Q) and also know that P, you can conclude Q, as we did in the above example. This is practically the same as what we called *modus ponens* in Chapter 5. It is a very well-known, simple, intuitive, and obviously valid form of reasoning. So it deserves its place as one of our basic rules. It is shown below.

	⋮	
i.	(P ⊃ Q)	
	⋮	
j.	P	
	⋮	
∴	Q	i, j, ⊃E

The Rule of Conditional Elimination

It is important that you interpret this rule correctly. Even though it is called Conditional *Elimination*, it does *not* allow you to simply *eliminate* or get rid of the horseshoe whenever you want. This might very well be the commonest mistake of all among logic students (to derive Q from (P ⊃ Q) alone). It is intuitively very clear to all rational people that just because you know (P ⊃ Q), does not mean you can conclude Q or that you can conclude P. Suppose you believe that if Smith accepts bribes then he is corrupted. Clearly, that is not to say that Smith *is* corrupted. All we know is that he is corrupted *in case* he accepts bribes. So until we know, from another source, that he did accept bribes, we cannot conclude anything about his corruption. Similarly, from the information that "If Smith accepted bribes, then he is corrupted", you cannot conclude that he did accept bribes. Even though these facts are intuitively very obvious to most people, when it comes to doing proofs with symbols, for some reason students tend to take either Q or P out of (P ⊃ Q) without citing any other lines. Watch out for these mistakes. One way to avoid them is to always remember that, as shown above, Conditional Elimination requires you to cite *two* line numbers. One line has to have a conditional claim in it, and the other has to have the antecedent of that conditional.

1.	(P ⊃ Q)	A
2.	P	A
∴	Q	1, ⊃E

Incorrect application of ⊃E: Two line numbers needed

1.	(P ⊃ Q)	A
∴	P	1, ⊃E

Incorrect application of ⊃E: Two line numbers needed

Here is a proof that we can build, using Conjunction Elimination and Conditional Elimination.

1.	(P ⊃ E)	A
2.	(P . R)	A
3.	P	2, .E
∴	E	1, 3, ⊃E

We needed P to derive E, so first we got P from line 2 and then used Conditional Elimination to get E. We did not need the sentence R in order to derive our desired conclusion.

Here is another proof that uses Conjunction rules and Conditional Elimination.

1.	(A ⊃ (B . ~C))	A
2.	A	A
3.	(B . ~C)	1, 2, ⊃E
4.	B	3, .E
5.	~C	3, .E
∴	(~C . B)	4, 5, .I

First we got (B . ~C) from 1 and 2, using horseshoe elimination, then switched the two sentences around, using our two conjunction rules.

§6.6 Conditional Introduction

As I said above, the elimination rules have the connective in the premises but not necessarily in the conclusion. The introduction rules, on the other hand, have the connective in their conclusion but not necessarily in their premises. So, a Conditional Introduction rule would be a rule that allows you to put a horseshoe *in the conclusion* without having it in the premises. To see how that works, suppose you want to prove "If Alex hits another homerun this season, then he will retire". This is a conditional claim. Proving it must be very different from proving that, say, Alex does hit another homerun or that he will retire, because these are not conditional claims. Here is one way you might prove that if Alex hits another homerun this season, then he will retire:

> Suppose Alex hits another homerun this season. Given that this will be a new record, he will get a new RBI and will be the best active baseball player of all. But if he is all that, he will be at the peak of his career. But he can stay at the peak only if he retires now. Therefore, if Alex hits another homerun, then he will retire.

The strategy of the above argument is to *assume* that Alex scores a homerun and see what follows from that. We are not saying that he will or that he will not. We are simply assuming it for the moment, or as logicians say, *for the sake of the argument*, without committing ourselves to it one way or another. The rest of the argument derives a series of conclusions from that assumption and ends with the desired conclusion, namely, that Alex will retire. Finally, we have shown that *if* Alex scores a homerun, then he will retire.

So, if our desired conclusion is a conditional claim, we need to assume the antecedent for the sake of argument and see what follows from that. If we can show that the consequent follows under the assumption of the antecedent, we have proved our conditional conclusion. This is basically the content of the Conditional Introduction rule.

	⋮	
i.	P	A (for ⊃I)
	⋮	
j.	Q	
∴	(P ⊃ Q)	i - j, ⊃I

The Rule of Conditional Introduction

As you see, to use Conditional Introduction, you need to have a little proof *within* the original proof. This is called a *sub-proof*. Your sub-proof starts with an assumption, which must be your antecedent, and ends with a result, which must be your consequent. You then write your desired conditional conclusion *outside* the sub-proof. The justification will be i - j, ⊃I, which we read as: “i *through* j, and horseshoe introduction”. Deciding Alex’s career is left as an exercise for you.

Here is another derivation that uses Conditional Introduction.

1.	(F ⊃ G)	A
2.	(F . H)	A (for ⊃I)
3.	F	2, .E
4.	G	1, 3, ⊃E
∴	((F . H) ⊃ G)	2-4, ⊃I

The goal is to derive ((F . H) ⊃ G) from (F ⊃ G). If F is sufficient for G, then F and H together are certainly sufficient for G, and the above derivation proves that. Notice that we assumed (F . H) because that is the antecedent of our *conclusion*. You need to look at your *conclusion* to find out what you need to assume. Do not look at your premises for that. If you look at the premises, you might think that you need to assume F. That would not help you get your conclusion.

Once you learn that you're allowed to assume stuff, it is tempting to assume whatever you feel like you need for the proof. But be careful, you shouldn't assume things unless for a rule that *requires* sub-proofs. (So far, the only rule that requires a sub-proof is Conditional Introduction.) So every time you assume something, you need to know *for the purpose of which rule* you did that. That is why we write "for ⊃I" in parentheses in front of our assumptions when we use ⊃I. For instance, the proof below is incorrect.

1.	(F ⊃ G)	A
2.	(F . H)	A
3.	| F	A (for ⊃E)
∴	| G	1, 3, ⊃E

The conclusion cannot be inside a sub-proof

It is incorrect for a number of reasons. For one thing, ⊃E does not work with sub-proofs. You can never assume something and justify it by "A (for ⊃E)". Secondly, the conclusion of an argument can never be inside a sub-proof. A sub-proof is part of the intermediary steps that you take to reach a conclusion. The conclusion itself must be outside the sub-proof.

Neither of these two problems is correctable. Suppose, for example, that you change "for ⊃E" to "for ⊃I". That wouldn't work, because then all that you can possibly conclude from your sub-proofs is (F ⊃ G), in accordance with Conditional Introduction. That is not your desired conclusion: you want to derive G! Now suppose someone tries to fix the second problem by reiterating G outside the sub-proof as shown on the next page. This would also be incorrect because once a sub-proof line is terminated, the sentences inside are no longer available. You cannot reiterate a line from inside a sub-proof, unless the sub-proof is not closed yet.

1.	(F ⊃ G)	A
2.	(F . H)	A
3.		F	A (for ⊃E)
4.		G	1, 3, ⊃E
∴	G	4, R

Incorrect application of R: Once a sub-proof is closed, the lines inside it are no longer available

The following rule helps you remember not to do as above:

Once a sub-proof line ends, we say that that sub-proof is "terminated" or "closed", and the sentences proved inside that sub-proof are no longer available for future use.

We need to follow this rule or else we will be able to prove virtually anything from anything.

It is possible that we have a sub-proof within a sub-proof within the original proof. In general, there is no limit on the number of sub-proofs that can go within other (sub-)proofs. Study the following proof carefully:

1.	(A ⊃ B)	A
2.	((B . ~D) ⊃ E)	A
3.		A	A (for ⊃I)
4.		B	1, 3, ⊃E
5.			~D	A (for ⊃I)
6.			(B . ~D)	4, 5, .I
7.			E	2, 6, ⊃E
8.		(~D ⊃ E)	5-7, ⊃I
∴	(A ⊃ (~D ⊃ E))	3-8, ⊃I

The reason we need two sub-proofs, one within the other, is that our conclusion has two horseshoes in it, one within the scope of the other. To prove a horseshoe sentence, we assume its antecedent (in this case A) and derive its consequent (in this case (~D ⊃ E)). But the consequent itself is another horseshoe claim. Thus, we need to set up another sub-proof for that one.

§6.7 Negation Elimination and Introduction

Sometimes, whether in everyday life, or in mathematics, science, or philosophy, you cannot directly prove that your desired conclusion is true. Instead, what you can do is show that it is not false. In other words, sometimes you might not be able to prove, using direct methods, that some statement is the case, but you can prove it by showing that the opposite statement cannot possibly be the case.

Consider, for example, atheism. An atheist needs to prove that there is no God. (In Chapter 2 we saw that some atheists use abductive reasoning, based on the criterion of simplicity, but here we want to know if any *deductive* argument for atheism is possible.) It is not easy to think of a direct argument that starts from some facts in the premise and ends with the conclusion "There is no God". But some atheists have made the following, indirect argument:

> Let's suppose there is a God. Then, presumably, he is benevolent, all-knowing, and all-powerful. A being with all these properties can easily prevent any evil or harm to his creation, because 1) provided he is benevolent, he *wants* to prevent evil and harm, 2) provided he is omniscience, he *knows* when and how the evil or harm is going to occur as well as how to prevent it, and 3) provided he is omnipotence, he *is able to* prevent the evil or harm, if he wants to. That a being wants, knows how, and is able to prevent evil or harm is sufficient for there being no evil or harm. Therefore, there must be no evil or harm in the world. But there is evil and harm everywhere in this world (e.g. birth defects, child hunger, world poverty, earthquakes, volcanoes, etc.). Therefore, there is no God.

The argument is rather long, but the strategy is clear. We assume, for the sake of argument, that there is a God. Then we show that a contradictory result follows from that assumption, namely that there is evil and harm and there is no evil and harm in the world. Since contradictions are unacceptable, we then reject the original assumption. This argumentative strategy has also been known since antiquity and has its own Latin name. It is called *reductio ad absurdum*, which means "reducing to absurdity". We assume something in order to reduce it to absurdity, the absurdity being a contradiction. Therefore, we have shown that the assumption cannot possibly be true. (By the way, you might find that you disagree with the above argument against the existence of God. Since the argument above is valid, if you disagree with the conclusion, it must be because you think one of the premises is false.)

Negation Elimination and Negation Introduction are in a sense one and the same rule. They both correspond to *reductio ad absurdum*. Here is Negation Elimination:

```
   |________
   | .
   | .
   | .
i. | | ~P          A (for ~E)
   | |______
   | | .
   | | .
   | | .
   | | Q
j. | | ~Q
∴  | P             i - j, ~E
```

The Rule of Negation Elimination

So, the rule says, if you assume that ~P and show that this leads to a contradiction (namely Q and ~Q), you can reject the assumption of ~P by concluding that P.

Negation Introduction is quite similar, except for where the negations go. Here it is:

```
   |________
   | .
   | .
   | .
i. | | P           A (for ~I)
   | |______
   | | .
   | | .
   | | .
   | | Q
j. | | ~Q
∴  | ~P            i - j, ~I
```

The Rule of Negation Introduction

In this case, too, we assumed something and showed that it leads to contradiction, in order to reject it and conclude the opposite. But in Negation Elimination, your assumption has negation in it, while your conclusion need not. In the case of Negation Introduction, it's the other way around; your conclusion has negation in it, while the assumption need not. It is left as exercise to prove the above argument about God and evil using Negation Introduction.

Let's look at an example. At the beginning of this chapter, I asked you to think through a *modus tollens*. You can now see that the reasoning we used to prove that argument was also a Negation Introduction. We can symbolize it as follows:

1.	(P ⊃ Q)	A
2.	~Q	A
3.	| P	A (for ~I)
4.	| Q	1, 3, ⊃E
5.	| ~Q	2, R
∴	~P	3-5, ~I

First we assumed P for the purpose of Negation Introduction. Then we showed that Q follows based on premise 1 and our assumption of P. However, we know ~Q from premise 2. Thus, assuming P results in a contradiction. Therefore, we can reject that assumption and conclude that ~P. Incidentally, this was the first time that we *had to* use Reiteration. The above derivation cannot be done without Reiteration.

To have an example for Negation *Elimination*, let's see how we can prove (P ⊃ Q) from the assumption of (~Q ⊃ ~P). We must be able to prove it, because we know from the Law of Contraposition that they are equivalent. Two equivalent sentences can be proved from the assumption of one another. For pedagogical reasons, I'll go through the construction of this proof step by step. To start, we look at our conclusion. Since it is a conditional claim, we *guess* that the rule being used is Conditional Introduction. So we set up a ⊃I with P as its assumption and Q as its conclusion (see next page). Now our new goal is to get Q. But how do we get Q from the given premises and the assumption of P? This might seem strange because Q does not occur anywhere in the premises. The closest thing to Q that we have is ~Q in the first premise, but that hardly helps. In situations like this, when you realize that the sentence you're looking for is not to be found anywhere in your premises, or all that you can find is its negation, you should immediately think of negation rules (either Negation Elimination or Introduction). In this case, it would be Negation Elimination because our desired conclusion is Q, which does not have negation in it.

```
1. | (~Q ⊃ ~P)          A
   |-----------
2. | | P                A (for ⊃I)
   | |----
   | |
?  | | Q                ??
∴  | (P ⊃ Q)            2-?, ⊃I
```

So let’s set up a Negation Elimination within our sub-proof. We assume ~Q for the sake of argument, and hope to show a contradiction follows from that.

```
1.  | (~Q ⊃ ~P)                 A
    |-------------
2.  | | P                       A (for ⊃I)
    | |----
3.  | | | ~Q                    A (for ~E)
    | | |----
    | | |
    | | | (some sentence)
??  | | | ~(the same sentence)
?   | | Q                       3-??, ~E
∴   | (P ⊃ Q)                   2-?, ⊃I
```

But now it is easy to see how the contradiction comes about. Since we assumed ~Q, we can get ~P using line 1 and our assumption of ~Q. But we are also assuming P. So we have P and ~P at the same time and therefore the assumption of ~Q has to be rejected. You can see the completed proof on the next page.

1.	(~Q ⊃ ~P)	A
2.	| P	A (for ⊃I)
3.	| | ~Q	A (for ~E)
4.	| | ~P	1, 3, ⊃E
5.	| | P	2, R
6.	| Q	3-5, ~E
∴	(P ⊃ Q)	2-6, ⊃I

Once again, remember that it does not help to assume stuff at random. Make assumptions (sub-proofs) only if you have good reasons for doing so!

§6.8 Disjunction Elimination

Let's turn to disjunction rules, and first disjunction elimination. Remember that when we call a rule an "elimination" rule, it is because our premise has that connective in it, but our conclusion need not. So you might think that a disjunction elimination rule will somehow help us get rid of the disjunction. But one thing this rule can definitely *not* look like is this: "If you have (P ∨ Q) you can conclude P using disjunction elimination". This would be an invalid argument. If all you know is that it is either going to rain or hail, you cannot conclude that it is going to rain. Knowing that either of two things is going to happen is less information than knowing that one of them definitely happens. And it is always invalid to infer more from less (remember Chapter 1?). So, there will be no rule that allows us to simply get rid of the disjunction.

But the following argument *is* valid:

It is going to either rain or hail.

If it rains I'll need an umbrella, and if it hails I'll need an umbrella.

Therefore, I'll need an umbrella.

In other words, if you know that either of two things is going to happen, and also that either way some conclusion can be made, then you don't have to wait to see which of the two possibilities is going to occur. You can derive your conclusion already.

Disjunction Elimination exploits the above fact. Here is the rule:

```
    |________
    | :
i.  | (P ∨ Q)
    | :
j.  | | P          A (for ∨E)
    | |______
    | | :
k.  | | R
    |
l.  | | Q          A (for ∨E)
    | |______
    | | :
m.  | | R
∴   | R            i, j - k, l - m, ∨E
```

The Rule of Disjunction Elimination

The rule looks long and complicated, but don't be intimidated. It merely states that if you know for a fact that either P or Q is going to happen, and in addition can show that if P happens R follows and if Q happens also R follows, then you have shown that R is the case. The justification must cite the line in which the disjunction occurs and the two sub-proofs, as well as the rule of $\vee$E. So, $\vee$E is the only rule that requires *two* sub-proofs. Remember, again, that even though we call it Disjunction Elimination, it does *not* allow you to simply eliminate the disjunction (and, say, get P out of $(P \vee Q)$). You must follow the above pattern.

So let's see how our rain/hail example works out. The premises are the following. "It is going to either rain or hail", which would be $(R \vee S)$. "If it rains I need an umbrella", which is $(R \supset U)$. And, finally, "If it hails I need an umbrella", i.e. $(S \supset U)$. The conclusion is "I need an umbrella". The proof is shown below.

1.	$(R \vee S)$	A
2.	$(R \supset U)$	A
3.	$(S \supset U)$	A
4.		R	A (for $\vee$E)
5.		U	2, 4, $\supset$E
6.		S	A (for $\vee$E)
7.		U	3, 6, $\supset$E
∴	U	1, 4-5, 6-7, $\vee$E

This is the simplest proof that can be done with Disjunction Elimination. We will see more interesting ones later on.

The ultimate sign that you must use $\vee$E is that one of your assumptions has a disjunction claim in it that does not occur in that exact form elsewhere in the proof.

§6.9 Disjunction Introduction

Unlike the elimination rule for disjunction, Disjunction Introduction is a very simple rule. It simply says that you can always add, using disjunction, whatever sentence that you wish to the sentence you have. To understand this, it helps to imagine a situation in which you don't want to lie but don't want to tell the whole truth either. So, let's say you are captured by the enemy and are being interrogated about the location of some very important documents. You know that the documents are in the Major's drawer. But you are very ethical and have sworn never to lie under any circumstances. So, instead of telling them the whole truth and nothing but the truth, you tell them that the documents are either in the Major's drawer or in the Colonel's office, hoping that the additional information will distract them for a while. You haven't lied; what you said was true. So the following argument is valid:

The documents are in the Major's drawer.

Therefore, the documents are either in the Major's drawer or in the Colonel's office.

Remember that validity means the truth of the premises guarantees the truth of the conclusion, which is obviously the case about the above argument.

Disjunction Introduction is based on the above fact. It allows you to derive $(P \vee Q)$ from P. You can also derive $(Q \vee P)$ from P, if you want. Here is the rule:

	⋮	
i.	P	
	⋮	
∴	$(P \vee Q)$	i, ∨I

The Rule of Disjunction Introduction

	⋮	
i.	P	
	⋮	
∴	$(Q \vee P)$	i, ∨I

The Rule of Disjunction Introduction

The sentence that you add can be anything. In an actual proof, you know what sentence you need to add, based on your desired conclusion. Notice that, unlike Conjunction Introduction, Disjunction Introduction does not "put together" two sentences. Thus, the following is an incorrect application of ∨I:

1.	P	A
2.	Q	A
∴	$(P \vee Q)$	1, 2, ∨I

Incorrect application of ∨I:
No need to cite two line numbers

A correct application of ∨I only cites one line. In the above example, either line 1 or line 2 would be enough. Citing too many lines makes your proof wrong as much as citing too few lines does. The above proof would be correct if only line 1 or only line 2 were cited.

The combination of ∨E and ∨I allows us to derive (B ∨ A) from (A ∨ B). But the way this is done is quite different from how we derived (B . A) from (A . B). This time we can't "break it apart and put it back together" because there is no rule that allows us to break apart an or-claim. However, we know we need to use Disjunction Elimination, because our premise is a disjunction claim, namely (A ∨ B), which is nowhere else to be found in the proof. We said at the end of the last section that this is the ultimate sign that you must use Disjunction Elimination.

1.	(A ∨ B)	A
2.	A	A (for ∨E)
3.	(B ∨ A)	2, ∨I
4.	B	A (for ∨E)
5.	(B ∨ A)	2, ∨I
∴	(B ∨ A)	1, 2-3, 4-5, ∨E

Again, the idea is that we assume both sides of the disjunction, (A ∨ B), and show that either way our desired conclusion, (B ∨ A), follows. Thus, we can derive our conclusion by citing the disjunction, the two sub-proofs, and the rule of Disjunction Elimination. Note that we used ∨E because there was a disjunction in our *assumption*, and *not* because there was a disjunction in our conclusion.

This completes our list of rules. We promised to introduce two rules per connective plus Reiteration, and we decided to skip Biconditional Elimination and Biconditional Introduction.

§6.10 Strategies for Solving Proof Problems

There are two general strategies for solving proof exercises, which we shall call *the top-down* and *the bottom-up approach*. In the top-down approach we turn to our premises to give us clues about how to do the proof, whereas in the bottom-up approach we let our desired *conclusion* guide us. Neither approach is complete by itself, so you need to learn to combine them. But even the two approaches combined will not automatically tell you what to do; you need to use your own understanding and experience. Not to mention that these approaches can occasionally mislead you.

§6.10.1 The Top-down Approach

Roughly, the top-down approach suggests that you look at the *main connectives* of your assumptions. If the main connective is conjunction, conditional, or disjunction, you are advised to think about Conjunction Elimination, Conditional Elimination, or Disjunction Elimination, respectively. The top-down approach does *not* apply to negation, and does *not* say anything about any of the introduction rules.

The most clear-cut application of the top-down approach is when there is a disjunction in one of your premises. In such a case, there is a very good chance that you need to immediately set up two sub-proofs for ∨E. We saw two examples of this above. Let's work out another example here. Consider this argument: (S ⊃ ~Z), (S ∨ (B . ~Z)), ∴ ~Z. Following the top-down approach, we examine our premises first. The first premise has conditional as its main connective. Thus, we know that we will probably need to use Conditional Elimination later on. The second premise has disjunction as its main connective, which indicates that Disjunction Elimination is needed. When you have two candidate rules and one of them is Disjunction Elimination, always let ∨E take over. So, here is the setup:

1.	(S ⊃ ~Z)	A
2.	(S ∨ (B . ~Z))	A
3.		S	A (for ∨E)
		~Z	
		(B . ~Z)	A (for ∨E)
		~Z	
∴	~Z	2, 3-?, ?-?, ∨E

We have assumed each of the two sides of the disjunction separately, and set out to prove our conclusion once in each sub-proof. Notice that so far we have done nothing but mechanically set up what the rule requires. We did not use any special skills or make any smart moves. Nonetheless, we are almost finished! So following the rules can greatly reduce the amount of work you need to do. Anyway, the rest is not hard to guess. In the first sub-proof, ~Z comes from lines 1 and 3, and in the second sub-proof it comes from the assumption of (B . ~Z). The full proof will therefore be as follows:

1.	(S ⊃ ~Z)	A
2.	(S ∨ (B . ~Z))	A
3.	S	A (for ∨E)
4.	~Z	1, 3, ⊃E
5.	(B . ~Z)	A (for ∨E)
6.	~Z	5, .E
∴	~Z	2, 3-4, 5-6, ∨E

As another example of the top-down approach, let's consider this argument: (F . G), (F . H), ∴ ((G . H) ∨ (R . H)). First look at the first premise. It has a conjunction as its main connective, so the top-down approach tells us that there is a good chance we need to use Conjunction Elimination. This is the rule that allows us to break the sentence apart. So, the top-down approach suggests that we take either F or G or perhaps both out of the first premise for future use. If you look at your conclusion, F does not occur in it, but G does. So perhaps we need to separate G from premise 1. Also, applying the same reasoning to the second premise, we can make the guess that H needs to be separated from it for future use. Thus, so far we have the following on our scrap paper:

1.	(F . G)	A
2.	(F . H)	A
3.	G	1, .E
4.	H	2, .E
∴	((G . H) ∨ (R . H))	

Now it is not difficult to see where this is going. We can just put G and H together using Conjunction Introduction and then use Disjunction Introduction to add the extra sentence to it which gets us the desire conclusion.

1.	$(F \cdot G)$	A
2.	$(F \cdot H)$	A
3.	G	1, $\cdot$E
4.	H	2, $\cdot$E
5.	$(G \cdot H)$	3, 4, $\cdot$I
$\therefore$	$((G \cdot H) \vee (R \cdot H))$	5, $\vee$I

Note that we did *not* use the top-down approach (or any "approach") for step 5 and the last step. We just used our intelligence and understanding.

But don't apply the top-down approach mechanically. The approach merely provides you with a guess. There is no guarantee that you will be using, say, Conjunction Elimination just because one of the premises has Conjunction in it. Consider this argument: $(F \cdot G)$, $((F \cdot G) \supset J)$, $\therefore$ J. If you focus on the first premise, you might be misled into thinking that Conjunction Elimination is required to break it apart. But if you glance at the second premise, you see that the exact same sentence, $(F \cdot G)$, shows up there too, so there is no need to break it apart. Simply use 1, 2, $\supset$E to get J.

Another shortcoming of the top-down approach is that it merely helps you guess what rule you will probably use, but it does not tell you *where* in the derivation and in what manner to use it. To see what I mean, let's look back at one of our previous examples: $(\sim Q \supset \sim P)$, $\therefore$ $(P \supset Q)$. The first premise has a horseshoe in it. So, according to the top-down approach, you will probably use Conditional Elimination somewhere in this proof. But where? Obviously, you can't use it immediately, because the rule requires that you have the antecedent as well, which you don't. We saw on p. 98 how Conditional Elimination was used on line 4. So the top-down approach tells you (probably) what, but cannot tell you where and how.

§6.10.2 The Bottom-up Approach

In the following, "desired conclusion" means either the final conclusion of the proof or any of the intermediary conclusions that we need in order to get the final conclusion.

In the bottom-up approach we let the desired conclusion guide us. Roughly, if the main connective of your conclusion is conjunction, conditional, or disjunction, you are advised to set up Conjunction Introduction, Conditional Introduction, or Disjunction Introduction, respectively. This will provide you with new desired conclusions. Apply the bottom-up approach to your new desired conclusions repeatedly, as long as you need. The bottom-up approach also says that if your desired conclusion is a simple sentence letter or a negated simple sentence letter, you need to look back at the previous lines (including the premises) to see if it can be found anywhere there. If you find it there, use the top-down approach and your own understanding to extract it from there. The bottom-up approach does *not* apply to negation and does not say anything about the elimination rules.

Let's see how all this works out through an example. We want to prove the following argument: (E ⊃ C), (H ⊃ ~Z), ∴ ((H • E) ⊃ (C • ~Z)). Following the bottom-up approach, since the main connective of the conclusion is horseshoe, we set up a Conditional Introduction: we assume the antecedent of the conclusion, and set out to derive the consequent. Don't let your premises trick you into thinking that you need to assume E or H. That won't get you anywhere. It is your *conclusion* that tells you what to assume. Here is how we set up the proof.

1.	(E ⊃ C)	A
2.	(H ⊃ ~Z)	A
3.	(H • E)	A (for ⊃I)
	(C • ~Z)	
∴	((H • E) ⊃ (C • ~Z))	⊃I

If you are good at proofs, you should already see the path. But let's see how the bottom-up approach guides us through it. Our new desired conclusion is now (C • ~Z). Focus on this new conclusion and try to forget about the rest for now. Applying the bottom-up approach again, we set up a Conjunction Introduction, because the main connective of the desired conclusion is conjunction. The "setup" looks like this:

1.	(E ⊃ C)	A
2.	(H ⊃ ~Z)	A
3.	(H . E)	A (for ⊃I)
	C	
	~Z	
	(C . ~Z)	.I
∴	((H . E) ⊃ (C . ~Z))	⊃I

You should understand the setup as saying "somehow, we need to get C by itself, and ~Z by itself so that we can put them together using .I and get (C . ~Z)". Now we have reached a point where our desired goals (namely C and ~Z) are either simple sentence letters (C) or negated simple sentence letters (~Z). The bottom-up approach says that at this point we need to look back at our premises or other available lines above to see if they can be found there. Obviously, in this case we can get C from line 1 and ~Z from line 2. Both moves can be made using Conditional Elimination. But if we want to derive C using Conditional Elimination, we have to prove E first, because that's our antecedent. Remember that ⊃E only works if you have the antecedent on another line. You can see that E can be derived from line 3, i.e. our assumption of (H . E). Similarly, to get ~Z, first we need to prove H. But H can be derived from line 3. So we can complete the proof as shown below. Don't forget to fill out the incomplete justifications.

1.	(E ⊃ C)	A
2.	(H ⊃ ~Z)	A
3.	(H . E)	A (for ⊃I)
4.	E	3, .E
5.	C	1, 4, ⊃E
6.	H	3, .E
7.	~Z	2, 6, ⊃E
8.	(C . ~Z)	5, 7, .I
∴	((H . E) ⊃ (C . ~Z))	3-8, ⊃I

Since I've said that neither the top-down nor the bottom-up approach applies to the negation rules, let me explain how we find out that we need to use the negation rules (I've said this above, too). Usually, you should start thinking about the negation rules if your conclusion is something that you cannot find anywhere in the premises or the other previous lines. Especially, consider negation rules if you cannot find the desired letter in the premises but you do find its negated counterpart there. So, if your desired conclusion is P and you cannot find P anywhere in the premises or if all that you find is ~P, then think about the negation rules.

To see what I mean, let's study an example in which the negation rules are used. Suppose you are asked to prove the following argument: (B . (~C ⊃ ~B)), ∴ (A ⊃ (B . C)). At first you might be surprised that this is a valid argument, because A does not even exist in the premises. But it is valid (use a truth table test for your peace of mind).

First, we look at the premise, in accordance with the top-down approach, and make the guess that we will be using Conjunction Elimination, and perhaps Conditional Elimination later on (because conjunction is the main connective of (B . (~C ⊃ ~B)) and conditional is the main connective of (~C ⊃ ~B)). Then we turn to the conclusion and the bottom-up approach. We are advised to set up a ⊃I.

1.	(B . (~C ⊃ ~B))	A
2.	| A	A (for ⊃I)
	|	
	| (B . C)	
∴	(A ⊃ (B . C))	⊃I

Our new desired goal is therefore (B . C). The bottom-up approach suggests us to try Conjunction Introduction, which requires that we derive B by itself and C by itself and put them together. So we set that up as shown on the next page.

1.	(B . (~C ⊃ ~B))	A
2.	A	A (for ⊃I)
	B	
	C	
	(B . C)	.I
∴	(A ⊃ (B . C))	⊃I

We have two new desired conclusions now, B and C. The bottom-up approach says that when our desired goal is a simple sentence letter, we need to search the previous lines for it. B can be found in line 1 and it is very easy to get it from there. So let's get that out of our way:

1.	(B . (~C ⊃ ~B))	A
2.	A	A (for ⊃I)
3.	B	1, .E
	C	
	(B . C)	.I
∴	(A ⊃ (B . C))	⊃I

This is where things get tricky. Our desired conclusion is C but it is nowhere to be found in the above lines. All that we can find is ~C in line 1 and even that is given as the antecedent of a conditional claim. There is no way to get the antecedent out of a conditional. So, as said above, in this kind of situation we should start thinking about negation rules. Maybe we can show that ~C leads to a contradiction and should therefore be rejected. But now it is not hard to see where that contradiction comes from. If we assume ~C, with a little work we can get ~B, but we have also derived B. That will be our contradiction. You can see how all this plays out below.

```
1. │ (B . (~C ⊃ ~B))          A
   ├──────────────
2. │ │ A                      A (for ⊃I)
   │ ├────
3. │ │ B                      1, .E
4. │ │ │ ~C                   A (for ~E)
   │ │ ├────
5. │ │ │ (~C ⊃ ~B)            1, .E
6. │ │ │ ~B                   4, 5, ⊃E
7. │ │ │ B                    3, R
8. │ │ C                      4-7, ~E
9. │ │ (B . C)                3, 8, .I
∴  │ (A ⊃ (B . C))            2-9, ⊃I
```

One more thing about the negation rules. The contradiction that follows under the sub-proof need not always be a simple sentence letter and its negation (such as B and ~B). It could be any sentence, simple or complicated, provided we can also prove its negation. Our contradiction could be, say, (P ⊃ ~Q) and ~(P ⊃ ~Q), or ((A ∨ ~B) ⊃ ~C) and ~((A ∨ ~B) ⊃ ~C), or whatever. So, unleash your imagination! An example is shown below.

```
1. │ ~(P ∨ ~Q)          A
   ├──────────
2. │ │ P                A (for ~I)
   │ ├────
3. │ │ (P ∨ ~Q)         2, ∨I
4. │ │ ~(P ∨ ~Q)        1, R
∴  │ ~P                 2-4, ~I
```

As you can see, the pair that serves as our contradiction here is (P ∨ ~Q) and ~(P ∨ ~Q). This point will come in handy in one of your homework questions below.

Thus we have seen how to apply the two approaches and how to combine them. The rest is purely a matter of exercise. Experience can teach you things that no amount of explanation can.

§6.E Homework

HW9

1- (1 pt. each) Prove the following arguments.

  a. (F • H), (F ⊃ ~G), ∴ (H • ~G)

  b. (~P ⊃ (Q • R)), ∴ (~P ⊃ (R • Q))

  c. (D ⊃ (~E • ~F)), (D • (G • K)), ∴ (G • ~E)

  d. (~R ⊃ S), (S ⊃ ~T), (S ⊃ U), ∴ (~R ⊃ (U • (S • ~T)))

  e. (P ⊃ ~Q), (R ⊃ S), ∴ ((P • R) ⊃ (~Q • S))

  f. (U ⊃ (V • T)), ∴ ((U ⊃ V) • (U ⊃ T))

2- (2 pts.) Consider the argument about Alex's baseball career again. It has the following premises. "Given that he hits another homerun this season, he will get a new RBI and will be the best active baseball player of all." "If he is the best active baseball player of all, he will be at the peak of his career." "He can stay at the peak only if he retires now." And the conclusion is: "If Alex hits another homerun, then he will retire.". Translate the premises and the conclusion into PL and prove it.

3- (2 pts.) In Chapter 5 we used truth tables to show that (P ⊃ (Q ⊃ R)) and ((P • Q) ⊃ R) are logically equivalent. Logical equivalence can also be shown by deriving each sentence from the other. So set up two proofs, one with (P ⊃ (Q ⊃ R)) as the premise and ((P • Q) ⊃ R) as the conclusion, the other with ((P • Q) ⊃ R) as the premise and (P ⊃ (Q ⊃ R)) as the conclusion. Complete both proofs, thereby showing that the two sentences are logically equivalent.

HW10

1- (1 pt. each) Prove the following arguments.

  a. (A • ~B), ∴ ((A ∨ C) • (~Q ∨ ~B))

  b. (~H ⊃ V), (~H ⊃ (~V • C)), ∴ H

c. ((~A ⊃ B) . ~B), ∴ ~~A

d. (~C ∨ D), (~C ⊃ D), ∴ D

e. (T ∨ S), (T ⊃ ~R), (~R ⊃ (S ∨ C)), ∴ (S ∨ C)

2- (2 pts.) Consider the atheist argument again. It has the following premises. “If there is a God, he is benevolent, all-knowing, and all-powerful”. “Provided he is benevolent, he *wants* to prevent evil”. “Provided he is all-knowing, he *knows* how to prevent evil”. “Provided he is all-powerful, he *is able to* prevent evil”. “That a being wants, knows how, and is able to prevent evil, is sufficient for there being no evil”. “There is evil and harm everywhere in this world”. And the conclusion is “There is no God”. Translate the argument into PL and prove it.

3- (1 pt.) Remember De Morgan’s first law: ~(P ∨ Q) = (~P . ~Q). As said above, two equivalent sentences can be derived from the assumption of one another. In this question we will derive (~P . ~Q) from the assumption of ~(P ∨ Q). First use the bottom-up approach. This will provide you with some new desired conclusions. Can your new desired conclusions be derived directly from the premises? If not, use negation rules to prove them. You’ll also need ∨I at some point.

4- (1 pt.) Let’s also try De Morgan’s second law ~(P . Q) = (~P ∨ ~Q). Again, given their equivalence, both sentences can be derived from the assumption of the other, but proving ~(P . Q) from the assumption of (~P ∨ ~Q) is easier than the other way around. To do so, first use the top-down approach. Set up the requirements of the rule carefully. Now you have to complete your sub-proofs. Your new desired conclusions can be proven using Negation Introduction.

5- (1 pt.) If a sentence is logically true, it can be proven with no premises. Prove ~(P . ~P) from an empty set of assumptions. This might seem strange at first, but given that your conclusion is nowhere to be found in the premises (there are no premises!) you know what rule to use.

Further Reading:

Merrie Bergmannn, James Moor, Jack Nelson (2009), *The Logic Book*, the McGraw-Hill Companies, Chapter 5

Teller, Paul (1989), *A Modern Formal Logic Primer*, available online, Volume I, Chapters 5 and 6 – Note: Teller uses a different rule of Disjunction Elimination. Everything else is pretty much the same.

Chapter 7

Quantificational Logic: Symbolism

§7.1 A Better Model of Language

In the last three chapters we saw how Propositional Logic models our deductive reasoning. That model is incomplete. While there are many arguments that it covers, there are others that it does not. In fact, some arguments that might come to mind as the first examples of valid reasoning cannot be represented in PL as being valid. Consider the very first valid argument that we introduced in this book (see Chapter 1).

All humans are mortal.

Socrates is a human.

Therefore, Socrates is mortal.

This is indeed a valid (and very intuitive) argument. What happens if we translate it into PL? There are no negations, conjunctions, disjunctions, conditionals, or biconditionals in any of the above sentences. Thus, the translation will simply be,

P

Q

R

which is clearly invalid. What this shows is that PL, as a model of language, is incapable of representing certain structures and it is exactly these structures that make the above argument valid. This is not surprising, because the smallest units in PL are sentences, so that all structure comes from connectives *between* sentences. But arguments such as the one above have their validity in virtue of the *internal* structure of the sentences.

We need a system of logic that can also represent the internal structure of sentences. Quantificational Logic (QL) makes a lot of progress in that direction. The good news is we can build QL on top of PL. In other words, we *enhance* our logic to include internal forms. In Chapter 1, we extracted the form of "All humans are mortal" as "All H are S". In QL, we will provide symbols to replace things like "all", "some", "none" and proper names like "Socrates".

QL is also a very precise language with strict rules and very few ingredients. This language is made of the following:

All ingredients of PL: Sentence letters, the five basic connectives, and parentheses, as well as all rules of PL are still in place.

Proper names: Lowercase letters of the alphabet, such as a, b, r, l, etc., represent names of individual people, places, objects, etc.

Predicates: Uppercase letters of the alphabet, such as F, G, C, R, etc. represent general categories, properties, or attributes of people and objects, when followed by lowercase letters.

Variables: The last four letters of the alphabet, namely w, x, y, and z, are placeholders for proper names. They help us express ideas such as "someone", "something", "anyone", and "anything".

Quantifiers: The symbol (∀x) means "For all x…" and the symbol (∃x) means "There is some x such that…". (∀x) is called the *universal quantifier* and (∃x) is called the *existential quantifier*. The parentheses come with the quantifiers and are not detachable.

The simplest sentence that we can build in QL consists of a predicate and a proper name. So the sentence "John is happy" will be translated as Hj. H stands for "is happy" and j stands for John. Logicians have decided to let *lowercase* letters stand for proper names, which is the opposite of the English conventional notation. So if you write HJ, it would not make any sense in QL. Likewise, hJ does not make sense. We must pick an uppercase letter for the property being attributed, in this case happiness, and a lowercase letter for the individual to which the property is being attributed, in this case John. The uppercase letter, which we call the *predicate,* comes first and the name is written immediately after that. "Eiffel is a tower" will be translated as Te. "Eiffel is not a tower" will be ~Te.

Now suppose we have the sentence "Something is burning". Since we don't know the object that is burning, we cannot write down a proper name for it. To express this, we write (∃x)Bx.

You should read this as “There is some x such that x is burning”. The string of symbols $(\forall x)Bx$ states that everything is burning. We read it formally as “For all x, x is burning”. The symbols Bx alone do not comprise a complete sentence. They correspond to “… is burning”, which is not a meaningful sentence of English. Here are some other examples of QL translations:

John is lucky but not happy. = $(Lj \cdot {\sim}Hj)$

Alex is German and Sara is American. = $(Ga \cdot As)$

Berlin is neither cold nor hot. = ${\sim}(Cb \vee Hb)$

Either Alex or Sara is American. = $(Aa \vee As)$

If the Earth is flat, then the Moon is made of cheese. = $(Fe \supset Cm)$

Everyone likes pizza. = $(\forall x)Px$

Someone doesn’t like pizza. = $(\exists z){\sim}Pz$

Everyone likes either pizza or spaghetti. = $(\forall w)(Pw \vee Sw)$

As you can see, it is entirely up to us whether we pick w, x, y, or z as our variable.

We can also express the idea of *relations* between objects in QL. So, to express the idea that John likes Mary, we write Ljm. In this case, L is called a *two-place predicate*, because it has two places for proper names. We can also have predicates with three or more places, but they do not concern us in this course. You can see some examples that use two-place predicates below:

John likes Mary but he doesn’t love her. = $(Ljm \cdot {\sim}Vjm)$

Someone likes Alex. = $(\exists x)Lxa$

Alex likes someone. = $(\exists x)Lax$

Everyone adores Mary. = $(\forall y)Aym$

Everyone likes Mary but Mary doesn’t like everyone. = $((\forall x)Lxm \cdot {\sim}(\forall x)Lmx)$

If Carol is smart then everyone loves her. = $(Sc \supset (\forall z)Vzc)$

Notice that the order of the lowercase letters is extremely important. “Someone likes Alex” and “Alex likes someone” are two very different sentences. It could be that one is true but the other is not.

A little explanation about the quantifiers is due. If you take a sentence such as (∀x)Hx, remove the quantifier, and replace the variable x with a proper name, you get what is called an *instantiation* of (∀x)Hx. So, Hm, Ht, Hb, etc. are all instantiations of (∀x)Hx. (But Hx is not an instantiation of (∀x)Hx, because x is not a proper name.) Similarly, Ha, Hh, etc. are also instantiations of (∃x)Hx. In the same manner, (Aa . As) is an instantiation of (∀x)(Ax . Ax) and so on. (∀x)Hx is to be understood as saying that *all* instantiations of the sentence are true. So Ha is true, Hb is true, and so on for every single object in the universe. (∃x)Hx, on the other hand, says that *at least one* instantiation of the sentence is true. In other words, either Ha is true, or Hb is true, or …. There is therefore some analogy between (∀x) and conjunction on the one hand, and (∃x) and disjunction (inclusive or) on the other hand.

Just like PL, QL has a *syntax*, i.e. a set of rules for building sentences. The sentences built in accordance with these rules are called *well-formed formulas*. Below is the syntax of QL:

Syntax of QL:

1- Every uppercase letter followed by a number of lowercase letters is a well-formed formula.

2- Every formula constructed from simpler well-formed formulas in accordance with the syntax of *Propositional Logic* is a well-formed formula.

3- Every quantifier followed by a well-formed formula is a well-formed formula, provided the formula involves the same variable as the quantifier.

4- Nothing else is a well-formed formula.

So, for example, (∀y)Gy is a well-formed formula, but (∀x)Gs or (∀x)Gy is not. The formula following the quantifier must have the same variables as the quantifier.

§7.2 Sentences with Multiple Predicates

Consider the sentence "All humans are mortal". How are we supposed to translate this into QL? We can't just write $(\forall x)Hx$ because that would mean "Everything (or everyone) is a human". $(\forall x)(Hx \cdot Mx)$ would also not work because this says "For all x, x is a human and mortal", which is not what the English sentence "All humans are mortal" says.

The correct translation of "All humans are mortal" in QL is this: $(\forall x)(Hx \supset Mx)$. This makes sense if you read it this way: "For all x, if x is a human then x is mortal". This sounds like just what our statement in English says.

Now consider the sentence "Some humans are mortal". One might be tempted to follow the same pattern and translate this as $(\exists x)(Hx \supset Mx)$. Unfortunately, this is not a good translation of our sentence. The reasons are complicated and go back to the truth table of the horseshoe. To cut the long story short, logicians have decided that the correct translation of "Some humans are mortal" is: $(\exists x)(Hx \cdot Mx)$. This can be read as "There is some x such that x is a human and mortal". This way of translating some-claims has certain ambiguities, but we shall not worry about them in this course.

Rule: Translate "All F's are G's" as $(\forall x)(Fx \supset Gx)$.

Rule: Translate "Some F's are G's" as $(\exists x)(Fx \cdot Gx)$.

Let's make our sentences one level more complicated. How should we translate "All small birds can fly"? Obviously, "All birds can fly" would be $(\forall x)(Bx \supset Fx)$. But what do we do with the predicate "small"? Comparing the two claims reveals the answer immediately. Just as "All birds can fly" says that being able to fly is conditioned upon being a bird, "All small birds can fly" says that being able to fly is conditioned upon being a bird *and* being small. So the right translation is: $(\forall x)((Sx \cdot Bx) \supset Fx)$. This says "For all x, if x is small and x is a bird, then x can fly". In analogy, "Some small birds can fly" must be translated as $(\exists x)((Sx \cdot Bx) \cdot Fx)$, which states that "There is some x such that x is small, x is a bird, and x can fly".

Rule: Translate "All F G's are H's" as $(\forall x)((Fx \cdot Gx) \supset Hx)$.

Rule: Translate "Some F G's are H's" as $(\exists x)((Fx \cdot Gx) \cdot Hx)$.

§7.3 The Universe of Discourse

If you want to make a long sequence of claims all of which are about, say, birds, it would be cumbersome to include the predicate "bird" in every sentence. Given that we are assuming all claims in this set to be about birds, making that explicit in every sentence would be unnecessary. Instead, we can simply specify it once at the beginning and drop B from all of our sentences. This is the idea behind the concept of a *universe of discourse*. For example, if you are told that the universe of discourse in a particular exercise is *everything*, and you are given the sentence "All birds can fly" to translate, you must translate it as $(\forall x)(Bx \supset Fx)$. But if you are told that the universe of discourse is *birds*, the same sentence must be translated as $(\forall x)Fx$.

> *The universe of discourse restricts the set of objects that our quantifiers range over. If your universe of discourse consists of objects of type T, don't assign a predicate for having the attribute T because that is already implicit in all sentences.*

The word "everyone" really means "everything that is a person". So if your universe of discourse is everything (or if it is not specified at all), we need to translate "everyone" and "someone" with an extra predicate for being a person. If, on the other hand, the universe of discourse is restricted to people, we drop the personhood predicate. See the following examples.

UD: Everything

Hx: x is happy

Gx: x is German

Rx: x is rich

Px: x is a person

Everyone is happy. = $(\forall x)(Px \supset Hx)$

Some Germans are rich. = $(\exists x)((Px \cdot Gx) \cdot Rx)$

All rich Germans are happy. = $(\forall x)((Px \cdot (Rx \cdot Gx)) \supset Hx)$

UD: People

Hx: x is happy

Gx: x is German

Rx: x is rich

Everyone is happy. = $(\forall x)(Hx)$

Some Germans are rich. = $(\exists x)(Gx \cdot Rx)$

All rich Germans are happy. = $(\forall x)((Rx \cdot Gx) \supset Hx)$

§7.4 The Big Four

There are four fundamental types of sentences that you must learn to translate. Let's call them "all-sentence", "some-sentence", "none-sentence", and "some-not-sentence". You already know all-sentences and some-sentences. Here are all four:

1) All A's are B's.
2) Some A's are B's.
3) No A's are B's.
4) Some A's are not B's.

The sentence "All A's are not B's" is ambiguous between 3 and 4. Stay away from it. If you ever find yourself inclined to utter this sentence, pause for a moment and ask yourself: Do I want to say that *none* of the A's are B? Or do I just want to say that not all of them are? If the former is what you mean, it is the same as 3, but if the latter is what you mean, it is the same as 4.

Let us first see what happens if the universe of discourse is all objects of type A. The translations for 1-4 above will then be, respectively,

UD: A-objects

1) All are B. = $(\forall x)Bx$
2) Some are B. = $(\exists x)Bx$

3) None are B. = ~(∃x)Bx

4) Some are not B. = (∃x)~Bx

These translations should be intuitively clear. But, interestingly, the third sentence can also be expressed using the universal quantifier. 3 says "It is not the case that there is some x such that x is B". In other words, if you examine the objects in our universe of discourse one by one, none is going to be B. But this is equivalent to "For all x, x is *not* B", because this also asks us to examine the objects one by one and claims that every one of them is going to be non-B. So, (∀x)~Bx is an equally good translation of 3.

Does 4 have any equivalents? "Some of them are not B" is the same as "Not all of them are B". (Hold on! Think about this for a moment.) Thus, another way of looking at 4 is that it simply denies "All of them are B". Consequently, ~(∀x)Bx will be an equally good translation of 4.

Rule: ~(∃x)Bx = (∀x)~Bx

Rule: ~(∀x)Bx = (∃x)~Bx

This should remind you of De Morgan's laws: the negation goes inside but changes ∃ to ∀ and vice versa. This is another way in which ∀ is similar to conjunction and ∃ to disjunction.

In the above, we had restricted the universe of discourse to A-objects. Let's relax that condition now and set the UD to everything. While the translations for 1 and 2 are still familiar from above, 3 and 4 are new. Below you see the translations for 1-4. Study them carefully.

UD: Everything

1) All A's are B's. = (∀x)(Ax ⊃ Bx)

2) Some A's are B's. = (∃x)(Ax . Bx)

3) No A's are B's. = ~(∃x)(Ax . Bx)

4) Some A's are not B's. = (∃x)(Ax . ~Bx)

We can find equivalents for 3 and 4 in this case too, but it is going to be slightly more complicated. One way to express the idea that “No A’s are B’s” is to deny that there is at least one A-object that is at the same time B. This is exactly what the above translation does. But another way to think about the same idea is this: Pick any object in the universe of discourse; if it is A, then it is not B. Therefore $(\forall x)(Ax \supset {\sim}Bx)$ is an equally good translation for 3. It guarantees that no A’s are going to be B’s.

Let’s think about 4 now. It states that there is at least one A-object that is not B. Another way to express the same idea is to say: It is not the case that all A’s are B’s (there are going to be some that are not). In other words, “Some Germans are not rich” is intuitively the same as “*Not all* Germans are rich”. This means that ${\sim}(\forall x)(Ax \supset Bx)$ is an equally good translation of 4.

Rule: ${\sim}(\exists x)(Ax \cdot Bx) = (\forall x)(Ax \supset {\sim}Bx)$

Rule: ${\sim}(\forall x)(Ax \supset Bx) = (\exists x)(Ax \cdot {\sim}Bx)$

You can base these rules on a firmer foundation by using some of the equivalence rules that we have learned so far. To get from ${\sim}(\exists x)(Ax \cdot Bx)$ to $(\forall x)(Ax \supset {\sim}Bx)$, first let the negation go inside and, of course, change the quantifier. So you’ll have $(\forall x){\sim}(Ax \cdot Bx)$. Then use De Morgan’s second law on this. Negation goes inside the parentheses and changes the “and” to “or”. We end up with: $(\forall x)({\sim}Ax \vee {\sim}Bx)$. Now use the equivalence rule that related disjunction to conditional (see Chapter 5). What you’ll find is exactly $(\forall x)(Ax \supset {\sim}Bx)$.

$${\sim}(\exists x)(Ax \cdot Bx) \xrightarrow{\text{negation goes inside}} (\forall x){\sim}(Ax \cdot Bx) \xrightarrow{\text{De Morgan's 2nd law}} (\forall x)({\sim}Ax \vee {\sim}Bx)$$

$$\xrightarrow[\text{and } \vee]{\text{Relation between } \supset} (\forall x)(Ax \supset {\sim}Bx)$$

You can apply a similar series of tricks to the second rule above, i.e. to get from ${\sim}(\forall x)(Ax \supset Bx)$ to $(\exists x)(Ax \cdot {\sim}Bx)$. This is left as an exercise for you.

§7.5 Complex Sentences and Sentences with Multiple Quantifiers

It happens quite often that two separate sentences of QL, each of which might have its own quantifier, are connected by a PL connective (conjunction, disjunction, etc.). To translate such sentences, first take care of the PL connectives and then translate each "chunk" using your QL skills. So, for example, suppose your UD is everything and your sentence is "If everyone is happy, no one should complain". This sentence has two chunks. Each chunk is a separate QL sentence. The translation will be ([something] $\supset$ [something]). The antecedent is "Everyone is happy" which is simply $(\forall x)(Px \supset Hx)$. The consequent is "No one should complain" which, as we saw above, must be translated as either $\sim(\exists x)(Px \cdot Cx)$ or $(\forall x)(Px \supset \sim Cx)$. Let's pick the first translation. Then the complete sentence, "If everyone is happy, no one should complain" will be $((\forall x)(Px \supset Hx) \supset \sim(\exists x)(Px \cdot Cx))$. Some further examples are given below. Study them carefully.

UD: Everything

If all birds can fly, then some mammals can't. = $((\forall x)(Bx \supset Fx) \supset (\exists y)(My \cdot \sim Fy))$

Some birds can't fly but some mammals can neither fly nor jump. = $((\exists z)(Bz \cdot \sim Fz) \cdot (\exists x)(Mx \cdot \sim(Fx \vee Jx)))$

Either no one is angry or all angry people are being polite. = $(\sim(\exists x)(Px \cdot Ax) \vee (\forall x)((Ax \cdot Px) \supset Lx))$

If everyone in the audience sings along, then I am happy. = $((\forall x)((Px \cdot Ax) \supset Sx) \supset H)$

It won't rain unless everyone dances Hula. = $(\sim R \vee (\forall x)(Px \supset Dx))$

In the last two examples, we used simple sentences letter of Propositional Logic in order to translate "I am happy" and "It will rain". Sometimes we borrow sentence letters from our PL toolbox. The sentences in question don't have straightforward quantificational translations.

Some of the above sentences had multiple quantifiers, but each chunk had only one quantifier. Multiple quantifiers can occur in a quite different way. We already know how to translate sentences such as "Everyone likes John" or "John likes everyone". But what if our sentence is "Everyone likes everyone"? We have two occurrences of "everyone" in this sentence, so it is natural to think that we'll need two universal quantifiers, like this: $(\forall x)(\forall y)Lxy$. Notice that we *cannot* translate it as $(\forall x)(\forall x)Lxx$. We can only have one quantifier per variable. Here the only variable is x, and repeating $(\forall x)$ would be incorrect. If we drop one of the quantifiers, on the other hand, we will get $(\forall x)Lxx$, which says that everyone likes themselves (For all x, x likes x, i.e. all objects in the universe of discourse like themselves)! Thus, the only correct translation of "Everyone likes everyone" is one with two *different* variables, such as x and y. Study the following sentences thoroughly:

UD: People

Everyone likes everyone. = $(\forall x)(\forall y)Lxy$

Someone likes someone. = $(\exists x)(\exists y)Lxy$

Everyone likes someone. = $(\forall x)(\exists y)Lxy$

Someone likes everyone. = $(\exists x)(\forall y)Lxy$

Someone is liked by everyone. = There is someone whom everyone likes. = $(\exists y)(\forall x)Lxy$

Everyone is liked by someone. = Everyone has someone who likes them. = $(\forall y)(\exists x)Lxy$

Everyone likes themselves. = $(\forall x)Lxx$

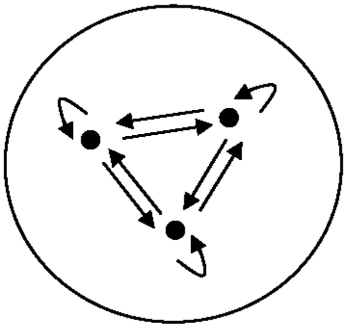
Someone likes themselves. = $(\exists x)Lxx$

Each of the sentences above has a different meaning from the others. No two of them say the same thing. Read each of them to yourself several times and make sure you understand, first, what it says and, second, why the correct translation is the one provided. It helps to always read $(\forall x)$ as "For all x, …" and $(\exists x)$ as "There is some x, such that …".

§7.6 Interpretations – A Very Quick Overview

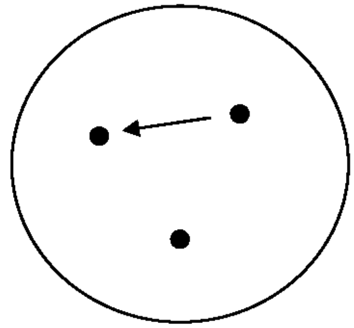
Diagrams can help visualize each of the above sentences. Below you see what are called *interpretations* of QL sentences. We simply specify a set of objects (concrete or abstract) that makes each sentence true. Thus, every interpretation in QL does roughly what a given row in a truth table did for us in PL (it describes a situation in which the sentence is either true or false). Normally, we would devote an entire chapter to interpretations, just as we did for truth tables in PL. But in this course we will merely have a quick look at some intuitive diagrams to help understand sentences with multiple quantifiers.

Below, we interpret a "person" as a little dot, and the relation of "like" as an arrow. An arrow pointing from a dot to another means the first "likes" the second. On the next page, you see one diagram for each of the sentences given above, except for the last two. For simplicity, all of these interpretations consist of universes that have only three objects in them. There is no specific reason behind choosing the number 3. Diagrams could be drawn with any number of dots.



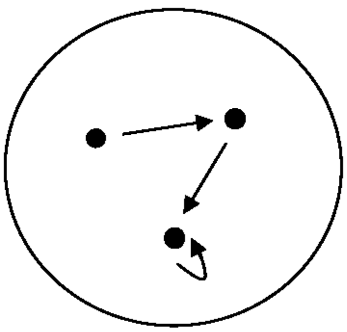
Everyone likes everyone

(∀x)(∀y)Lxy



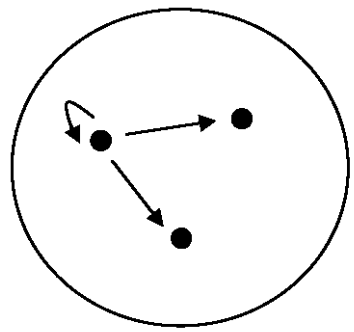
Someone likes someone

(∃x)(∃y)Lxy



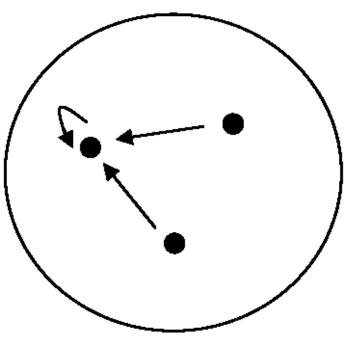
Everyone likes someone

(∀x)(∃y)Lxy



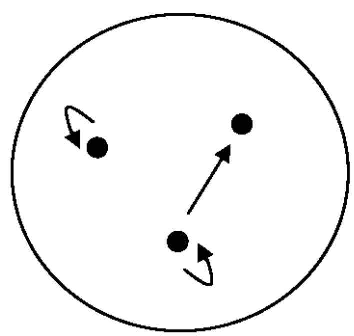
Someone likes everyone

(∃x)(∀y)Lxy



There is someone whom everyone likes

(∃y)(∀x)Lxy

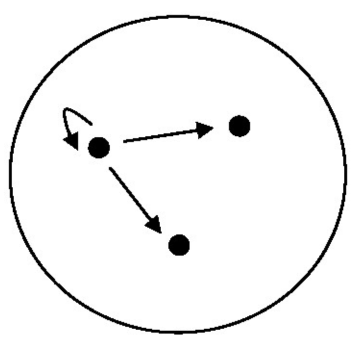


Everyone has someone who likes them

(∀y)(∃x)Lxy

One thing to notice about these sentences and diagrams is that in QL, the words "everyone" and "someone" always include oneself. So, for example, when we say "Everyone likes everyone", we are saying that everyone likes themselves *and* everyone else. Likewise, when we say "Someone likes everyone", that person who likes everyone must like themselves as well (that is, if our statement is true). Also, when we say "Everyone likes someone", that "someone" could be the person himself/herself, as in one of the diagrams below.

As I said, diagrams fulfill the role of truth tables for us, which means that we can use diagrams to test whether a given QL sentence is logically true/false, whether a given pair of QL sentences is logically equivalent, or whether a given argument in QL is valid. We shall not go through much detail here, but to see an example of how a diagram can establish that two sentences are inequivalent, let's compare "Someone likes everyone" and "There is someone whom everyone likes". The diagram below proves that these two sentences are logically inequivalent. Here is why. In this diagram, there is one dot that "likes" all other dots. However, there is no dot that all other dots "like": the first dot is only liked by itself, the second is only liked by the first, and the last dot is only liked by the first. So in that diagram it is not the case that there is someone whom *everyone* likes. This means that that diagram makes (∃x)(∀y)Lxy true but (∃y)(∀x)Lxy false. Therefore, these two sentences are not equivalent. You can apply the same concept to all other pairs of sentences from above and see that no two of them are equivalent. Notice that for any given pair of sentences, we only draw *one* diagram to prove inequivalence (the same way one row in the truth table would be enough to prove inequivalence). Sometimes students think that they can establish inequivalence by drawing two diagrams, one for each sentence, and point out that the two diagrams look different. If you draw two diagrams, each of which makes one of the two sentences true, you have not established anything at all. Inequivalence means that one sentence can be false in some situation in which the other sentence is true. It being true in some different situation does not prove anything.



Someone likes everyone: True

There is someone whom everyone likes: False

§7.E Homework

HW11

1- (5 pts.) Complete HET and HHT from LogiCola

Note: LogiCola uses (x) instead of (∀x).

2- (0.5 pt.) In §7.4 above, we used a series of equivalence "tricks" to turn ~(∃x)(Ax . Bx) into (∀x)(Ax ⊃ ~Bx). Apply similar tricks to convert ~(∀x)(Ax ⊃ Bx) into (∃x)(Ax . ~Bx) as follows. First, convert the horseshoe into a disjunction using the equivalence rule that related the conditional connective to disjunction (see Chapter 5). Second, let the negation in front of (∀x) go inside and make the appropriate changes. Then use De Morgan's first law to turn the disjunction into conjunction. Make sure your result is (∃x)(Ax . ~Bx).

3- (0.5 pt. each) Translate the following into QL. UD is *everything*.

a. New York is beautiful and everyone wants to move there.

b. Albert is not happy but he enjoys everything.

c. The Earth is neither flat nor round.

4- (0.5 pt. each) Translate the following into QL. UD is *people*.

a. Everyone is hated by someone.

b. Not everyone hates someone although someone hates everyone.

c. No one hates themselves. (Hint: What is this sentence the negation of?)

5- (0.25 pt. each) For each of the following sentences, draw a dot-arrow diagram that makes the sentence true (an arrow means "hates").

a. No one hates themselves.

b. Someone hates themselves.

c. Everyone is hated by someone.

d. Someone is hated by everyone.

6- (0.25 pt. each) Use dot-arrow diagrams to show that the following sentences are <u>not</u> logically equivalent.

a. "Everyone hates themselves" and "Everyone hates someone".

b. "Everyone is hated by someone" and "Everyone hates someone".

Further Reading:

Gensler, Harry J. (2002), *Introduction to Logic*, Routledge, Chapter 5, Sections 5.1 and 5.4

Teller, Paul (1989), *A Modern Formal Logic Primer*, available online, Volume II, Chapters 1 through 4

Merrie Bergmannn, James Moor, Jack Nelson (2009), *The Logic Book*, the McGraw-Hill Companies, Chapter 7

Chapter 8

Quantificational Logic: Proofs

§8.1 Proofs in QL

Now that we have enhanced our logical language and endowed it with the tools of Quantificational Logic, it is time to enhance our proof system as well. We would like to have a system in which to represent and prove arguments such as the Socrates argument. In our proof system for PL, we had two rules for each connective, one for introducing that connective and one for eliminating it. We will follow the same pattern of rules here. We shall have two rules for each quantifier, which adds up to four rules: Universal Elimination, Universal Introduction, Existential Elimination and Existential Introduction. (Note, however, that quantifiers are not really "connectives" proper. We will nevertheless have introduction and elimination rules for them.)

Since Universal Elimination and Existential Introduction are significantly simpler, we shall go a little out of order in discussing our four rules. We will talk about the easy rules first and then discuss the more difficult ones.

§8.2 Universal Elimination

If you know that everyone is kind, then you know that David is kind, Ashley is kind, and so on. So the following argument is obviously valid:

Everyone is kind.

Therefore, David is kind.

This is basically Universal Elimination. You can see the rule on the next page. It states that if you have proven or assumed (∀x)Fx somewhere in your proof, you can write down *any instantiation* of the sentence later on. For the justification you must cite the line in which the universal claim is given and the rule of ∀E.

	. . .	
i.	$(\forall x)Fx$	
	. . .	
∴	Fa	i, ∀E

The Rule of Universal Elimination

As an example, consider the following proof:

1.	$(\forall x)Fx$	A
2.	Fa	1, ∀E
3.	Fr	1, ∀E
∴	(Fa . Fr)	2, 3, .I

First we used Universal Elimination to get the instantiations Fa and Fr, and then put them together using Conjunction Introduction.

With Universal Elimination available to us, we can now prove the famous Socrates argument.

All humans are mortal.

Socrates is a human.

Therefore, Socrates is mortal.

The translation of this argument in QL will be: $(\forall x)(Hx \supset Mx)$, Hs, ∴ Ms. The proof is shown on the next page. Basically, we eliminate the universal to get a truly conditional claim. After that, there is nothing new about our proof. It is a simple PL-proof in which we use Conditional Elimination to get Ms.

1.	(∀x)(Hx ⊃ Mx)	A
2.	Hs	A
3.	(Hs ⊃ Ms)	1, ∀E
∴	Ms	2, 3, ⊃E

Of course, (Ha ⊃ Ma) and (Hb ⊃ Mb) and all other instantiations also follow from premise 1. But the instantiation that we *need* here is (Hs ⊃ Ms). This is a strategy that we will use a lot in our derivations. Usually we want to get rid of the quantifiers first, carry out the calculations with the derivation rules of PL, and then reintroduce quantifiers, if necessary, to resume the QL language. In this case, our conclusion did not have any quantifiers in it.

If you have multiple connectives in one sentence, you must eliminate them one by one. So, for instance, suppose you have (∀x)(∀y)Rxy as your premise and want to prove Rab. First you need to get rid of (∀x) and write (∀y)Ray. Your next task will then be to get rid of (∀y) to get Rab. The following proof demonstrates this:

1.	(∀x)(∀y)Rxy	A
2.	(∀y)Ray	1, ∀E
∴	Rab	2, ∀E

It would have been incorrect to derive Rab in one step and cite 1, ∀E. Do it one quantifier at a time and start with the outermost quantifier. The outermost quantifier is your "main connective" and must be taken care of first. You *cannot* get rid of (∀y) until after (∀x) is gone.

1.	(∀x)(∀y)Rxy	A	
∴	Rab	1, ∀E	**Incorrect application of ∀E: Must eliminate one at a time**

1.	(∀x)(∀y)Rxy	A	
∴	(∀x)Rxb	1, ∀E	**Incorrect application of ∀E: The outmost quantifier first**

§8.3 Existential Introduction

If it is the case that New York is beautiful, then it is the case that something is beautiful. In other words, that something is beautiful is less information than that New York is beautiful and, as we have said before, it is always valid to infer less from more.

New York is beautiful.

Therefore, something is beautiful.

The rule of Existential Introduction allows you to make the above move. Here it is:

	⋮	
i.	Fa	
	⋮	
∴	(∃x)Fx	i, ∃I

The Rule of Existential Introduction

So, with this rule, you prove or assume the *instantiation* and it is the existential claim that you derive. Consider the following proof:

1.	(∀x)Fx	A
2.	Fa	1, ∀E
∴	(∃x)Fx	2, ∃I

Let's suppose Fx stands for "x is funny". Then we have derived "Something is funny" from "Everything is funny". That makes perfect sense. If everything is funny, then certainly at least one thing is funny, assuming that the universe of discourse is not empty. We could not do the opposite. It would be *invalid* to derive "Everything is funny" from the sole assumption that "Something is funny".

Let's give our Socrates example a further twist. Suppose you have the same premises, "All humans are mortal" and "Socrates is a human", but want to prove that "Something is mortal". So the argument in question is: $(\forall x)(Hx \supset Mx)$, Hs, $\therefore$ $(\exists z)Mz$. The right strategy is to first derive the conclusion that "Socrates is mortal" and then introduce the existential. Here is the proof:

1.	$(\forall x)(Hx \supset Mx)$	A
2.	Hs	A
3.	$(Hs \supset Ms)$	1, $\forall$E
4.	Ms	2, 3, $\supset$E
$\therefore$	$(\exists z)Mz$	4, $\exists$I

What we said about taking it one quantifier at a time is also true when using Existential Introduction. Consider the following proof:

1.	$(\forall x)(\forall y)Lxy$	A
2.	$(\forall y)Lay$	1, $\forall$E
3.	Lab	2, $\forall$E
4.	$(\exists y)Lay$	3, $\exists$I
$\therefore$	$(\exists x)(\exists y)Lxy$	4, $\exists$I

We proved that "Someone likes someone" from the assumption that "Everyone likes everyone". The point to keep in mind is that the two existential quantifiers could not have been introduced at once. We can only introduce one in each line and we must do that in the order that we want them to be.

Rule: Whether you are introducing or eliminating quantifiers, do it one quantifier at a time and in the desired order.

§8.4 Universal Introduction

Suppose you have an argument whose conclusion is a universal claim, such as the following:

All birds can fly.

Therefore, all wingless birds can fly.

The above argument is certainly valid, because the conclusion is just a more limited version of the premise (it contains less information than the premise). If you think the conclusion is false it is because the premise is false. But *if* the premise were true, the conclusion would have to be true. Anyway, we would like to be able to prove this argument in a derivation.

What is the reasoning process that we go through to get to this conclusion? Here is one:

> Well, we know all birds can fly. So pick any random object which is both wingless and a bird. Since it is a bird, it must be able to fly. Now, the object we picked was not a special wingless bird, but a generic one. Therefore, *all* wingless birds can fly.

In the above reasoning, the two crucial moves are "pick any random object" and "since the object we picked was not special, but generic". It is often much easier to work with claims about individuals than general claims about unknown objects. Thus, we made a transition to an individual claim in order to work out its consequences and eventually convert it back to a universal claim at the end. This happens a lot in mathematical and scientific arguments.

The "pick any random object" move is just Universal Elimination. But we need a rule of inference that allows us to make the final move from the individual conclusion to a generalized claim. How can this be done? Obviously, you cannot conclude that "Everything is blue" merely based on the assumption that "The sky is blue". So, there will be no rule that simply allows us to derive (∀x)Fx from Fa. Nevertheless, logicians have realized that deriving (∀x)Fx from Fa *is* valid provided that the generalization is done correctly and the object is a truly generic, i.e.

arbitrary, object. We will see what these conditions amount to in a moment. But first, to summarize, here is the rule:

	⋮	
i.	Fa	
	⋮	
∴	(∀x)Fx	i, ∀I

The Rule of Universal Introduction

Provided that,

1) The letter being replaced (in this case "a") does not occur in the line being derived (in this case (∀x)Fx),
2) The letter being replaced (in this case "a") does not occur in any open assumption.

Of course, if your letter is something other than "a", *that* letter must satisfy 1 and 2 above. Let me explain condition 1 first. It says that "a" should not occur in the conclusion being derived. You might think that this condition is unnecessary because, obviously, "a" is nowhere to be found in the sentence (∀x)Fx. But if, instead of Fx, we had a two- or more-place predicate, it might have contained "a" in it. So, for example, if you have Laa on line i, condition 1 says that you cannot derive (∀x)Lxa from it, because "a" occurs in (∀x)Lxa. Consider, for example, the following proof.

1.	(∀x)Lxx	A
2.	Laa	1, ∀E
∴	(∀y)Lya	2, ∀I

Incorrect application of ∀I:
"a" occurs in the conclusion

The problem with this proof is that "a" occurs in the conclusion, thereby violating condition 1. When the universal quantifier has been introduced, only one of the a's has been replaced. This delegitimizes the proof. An example can clarify why such a move is invalid intuitively: We cannot conclude that everyone likes Alex (the conclusion) based on the fact that Alex likes himself (line 2 of the proof)!

Let's examine condition 2 now. The letter being replaced should not occur in any open assumption, or else the argument will be invalid. The following proof violates condition 2:

1.	(∀x)Fxb	A
2.	Fbb	1, ∀E
∴	(∀y)Fyy	2, ∀I

Incorrect application of ∀I:
"b" occurs in an open assumption

In this case, condition 1 is *not* violated because the letter "b" does not occur in the conclusion. But condition *2* is violated because "b" occurs in line 1, which is an open assumption. Note that the move from line 1 to line 2 is OK. It is valid to conclude that Alex likes himself based on the assumption that everyone likes Alex. But the move from 2 to 3 is incorrect, because it violates condition 2. Intuitively, the original premise only says that everyone likes Alex, but the conclusion says that everyone likes themselves, which is quite irrelevant to the premise.

The following proof does *not* violate either of the conditions:

1.	(∀x)Lxx	A
2.	Lcc	1, ∀E
∴	(∀y)Lyy	2, ∀I

You can also derive (∀x)Lxx from (∀y)Lyy in the same manner. This proves that (∀x)Lxx and (∀y)Lyy are equivalents, which is not surprising at all. In any case, condition 1 is not violated because (∀y)Lyy does not contain the letter "c". Condition 2 is also met because "c" does not occur in any open assumptions.

These two conditions might still be confusing, so let me explain them some more. Here is another way to understand condition 1: When universalizing a claim, replace *all* occurrences of the letter being removed, not just some. If you are removing the letter "a" to replace it with, say, "x", replace *all* a's with x's; don't leave any out. And here is another way to understand condition 2: the letter being removed and replaced with a variable must be an *arbitrary* letter. A given letter is arbitrary if and only if no previous assumptions about it have been made. This way we make sure that the proper name could be any old object's name and does not refer to any special individual. When an object's name occurs in an open assumption, that makes the object special and non-arbitrary.

Now we are ready to prove our argument about wingless birds given above. The premise is "All birds can fly", i.e. (∀z)(Bz ⊃ Fz), and the conclusion is "All wingless birds can fly", i.e. (∀z)((Wz . Bz) ⊃ Fz). The derivation is shown below. First, we get rid of the universal quantifier in the premise using ∀E. There is no restriction on this move. I have decided to use the proper name "d" for a change. But what do we do now? Using the idea explained above, if we can somehow prove an instantiation of the conclusion for an arbitrary letter, such as ((Wa . Ba) ⊃ Fa) or ((Wd . Bd) ⊃ Fd) or whatever, we can use Universal Introduction to get our desired conclusion, i.e. (∀z)((Wz . Bz) ⊃ Fz). Since we used "d" in line 2, let's stick to "d" as our arbitrary letter. So we set out to prove ((Wd . Bd) ⊃ Fd).

1.	(∀z)(Bz ⊃ Fz)	A
2.	(Bd ⊃ Fd)	1, ∀E
?	((Wd . Bd) ⊃ Fd)	
∴	(∀z)((Wz . Bz) ⊃ Fz)	?, ∀I

Now, the conclusion we are after, namely ((Wd . Bd) ⊃ Fd), is a conditional claim, so the bottom-up approach suggests that we set up a ⊃I. As always, what we assume in our sub-proof is entirely dictated by the antecedent of our desired *conclusion*. (We do *not* look at the premises for guidance as to what to assume.) Also, as you remember, the end result of our sub-proof has to be consequent of the conclusion (see the next page). Now it is not difficult to see the path, especially if you seek help from the top-down approach. Dismantle line 3 to get Bd out of it. Then use Bd and line 2 to get Fd. Study the completed proof as shown on the next page.

Every time you use ∀I, you must make sure you did not violate conditions 1 and 2. In our current exercise, condition 1 is not violated because "d" doesn't occur in the conclusion. Condition 2 is also not violated, because, even though "d" occurs in line 3 and line 3 *is* an assumption, it is not an *open* assumption: it was closed on line 5. "d" also occurs in line 2, and that line is open, but it is not an assumption. Line 2 is just a regular line and it is OK that "d" occurs there.

1.	$(\forall z)(Bz \supset Fz)$	A
2.	$(Bd \supset Fd)$	1, ∀E
3.	$(Wd \cdot Bd)$	A (for ⊃I)
4.	Bd	3, .E
5.	Fd	2, 4, ⊃E
6.	$((Wd \cdot Bd) \supset Fd)$	3-5, ⊃I
∴	$(\forall z)((Wz \cdot Bz) \supset Fz)$	6, ∀I

Notice that except for what we did in line 2 and in the last line, the rest has nothing to do with Quantificational Logic. Lines 3-6 were purely a matter of Propositional Logic.

§8.5 Existential Elimination

We have one more rule to go. This rule is to allow us to *eliminate* the existential quantifier from a sentence. Sometimes knowing that *something* or *someone* has a certain property is enough for drawing a conclusion. We might not need to know exactly *what* or *who*. Here is an example:

Whoever drinks this poison will die.

Someone has drunk the poison (say, because the bottle is empty).

Someone has died due to poisoning.

Premise 2 and the conclusion are both existential claims. The conclusion is no problem because we can use Existential Introduction to put the existential quantifier in it. But we still don't know how to deal with premises that have existential in them. Usually, when we reason through an argument such as above, we use a strategy similar to the "pick a random object" strategy that we talked about before. Consider the following reasoning:

> We know that someone has drunk the poison. Let's call that person Mr. or Ms. Anonymous. Since Anonymous drank the poison, he or she must be poisoned and will inevitably die soon. We don't know who he or she is ("Anonymous" was just an arbitrary name). Therefore, *someone* has died due to poisoning.

We need a rule that allows us to get rid of the existential (at least temporarily), because, as was said above, working with claims about individuals is much easier than working with claims about any or every object. The rule of Existential Elimination does exactly that.

```
   | 
   |______________
   | .
   | .
   | .
i. | (∃x)Fx
j. | | Fa            A (for ∃E)
   | |________
   | | .
   | | .
   | | .
k. | | R
∴  | R               i, j-k, ∃E
```

The Rule of Existential Elimination

Provided that,

1) The letter being replaced (in this case "a") does not occur in the sentence being derived (in this case R),
2) The letter being replaced (in this case "a") does not occur in any open assumption other than the one in the sub-proof (in this case Fa),
3) The letter being replaced (in this case "a") does not occur in the original existential claim (in this case (∃x)Fx).

The rule looks complicated, but it's not really different from the case of Mr. or Ms. Anonymous. We know that someone has drunk the poison. This is our (∃x)Fx. Then we assume that Anonymous is the one who did it. This is our assumption of Fa in the sub-proof. After a few steps, we prove that someone has died. This is our R. But conditions 1-3 make sure that nothing in our derivation depended on who in particular Anonymous is. Thus, our conclusion of R does not depend on our assumption of Fa and can therefore be pulled out of the sub-proof.

It helps to see some examples before discussing the reasons behind conditions 1-3. Here is our first example. We know that the sentences (∃x)Fx and (∃z)Fz mean the same thing, so we must be able to prove one from the assumption of the other. The following proof does that:

1.	(∃x)Fx	A
2.	Fa	A (for ∃E)
3.	(∃z)Fz	2, ∃I
∴	(∃z)Fz	1, 2-3, ∃E

Make sure you can follow the proof. First we assume an instantiation of our existential claim. I have picked the letter "a" for this instantiation, but it could have been any old letter. Then we prove that our desired conclusion follows under that assumption. In this case, our conclusion happens to be (∃z)Fz. So we prove that. The justification for this conclusion happens to be ∃I. Don't be confused by this. Generally, the use of Existential Introduction has nothing to do with the use of Existential Elimination. And the justification for the conclusion cites the ∃E rule, not because the conclusion has existential in it, but because the *premise* does.

We need to check conditions 1-3. Condition 1 is not violated in the above proof because the letter "a" does not occur in (∃z)Fz (obviously, of course!). Condition 2 is met because "a" does not occur in any open assumption other than line 2. And, finally, condition 3 is met because "a" does not occur in (∃x)Fx (again, obviously).

As another example, let's go back to our "poisonous" argument and work it out. The first premise is "Whoever drinks this poison will die", i.e. (∀x)(Px ⊃ Dx). The second premise is "Someone has drunk the poison", i.e. (∃x)Px. The conclusion, "Someone had died due to poisoning", can be translated as (∃x)(Px **.** Dx). Since one of our premises is an existential claim, we need to set up Existential Elimination, which requires a sub-proof. The assumption of the sub-proof must be an instantiation of our existential claim with an arbitrary letter, and the conclusion must be identical to the desired conclusion. So, we set up a sub-proof, with Pa as the assumption and (∃x)(Px **.** Dx) as the conclusion. This is shown on the next page. I would like to emphasize that the conclusion of the sub-proof for ∃E must be identical to your desired conclusion. Of course, the "desired conclusion" may or may not be the final conclusion of the proof. Sometimes you might need to use Existential Elimination to prove an intermediary conclusion, in which case *that* conclusion should be copied inside the sub-proof. Nevertheless, in most cases of concern to us, the conclusion in question is just the final conclusion. At any rate, in ∃E you will always have two identical sentences, one inside and one outside the sub-proof. If it helps, you can think of it as "pulling it out" of the sub-proof.

1.	(∀x)(Px ⊃ Dx)	A
2.	(∃x)Px	A
3.	| Pa	A (for ∃E)
?	| (∃x)(Px . Dx)	
∴	(∃x)(Px . Dx)	2, 3-?, ∃E

We can now see where this is going. We simply need to use ∀E on line 1 to get (Pa ⊃ Da) and combine that with line 3 to get Da. Then we can put Pa and Da together to get (Pa . Da) and use Existential Introduction to derive (∃x)(Px . Dx). Here it is:

1.	(∀x)(Px ⊃ Dx)	A
2.	(∃x)Px	A
3.	| Pa	A (for ∃E)
4.	| (Pa ⊃ Da)	1, ∀E
5.	| Da	3, 4, ⊃E
6.	| (Pa . Da)	3, 5, .I
7.	| (∃x)(Px . Dx)	6, ∃I
∴	(∃x)(Px . Dx)	2, 3-7, ∃E

Before we can rejoice, though, we need to make sure we didn't violate conditions 1-3. The letter "a" does not occur in (∃x)(Px . Dx) so condition 1 is satisfied. Nor does the letter "a" occur in any open assumption other than line 3, so condition 2 is satisfied as well. And, finally, the letter “a” does not occur in the original existential claim, namely (∃x)Px, in accordance with condition 3.

Now that we have seen how to apply ∃E, let’s inquire more deeply about the meaning of the rule. First the conditions. The explanations for these three conditions are virtually identical to the explanations we gave for conditions 1 and 2 of Universal Introduction. In a nutshell, conditions 2 and 3 make sure that our proper name is genuinely arbitrary and represents no special object, but a generic one. As we said above, for a proper name to be arbitrary, we must make sure we have made no previous assumptions about its features. If the letter does show up in earlier lines, it cannot hold the place for any random object, because it represents the specific object that those assumptions talk about. We want it to be a random instantiation, because we want to generalize over the result. Hence conditions 2 and 3. Condition 1, on the other hand, is there to make sure that the generalization is done correctly: Your conclusion must be a claim that doesn’t depend on the proper name being used. In particular, when generalizing a claim, replace *all* instances of the proper name with the variable, not just some.

It might serve as further explanation to note the close similarity between ∃E and ∨E. Recall Disjunction Elimination:

	⋮	
i.	(P ∨ Q)	
j.	| P	A (for ∨E)
	| ⋮	
k.	| R	
l.	| Q	A (for ∨E)
	| ⋮	
m.	| R	
∴	R	i, j - k, l - m, ∨E

The Rule of Disjunction Elimination

The idea behind ∨E was the following: Suppose you know that either P or Q happens. Suppose you can also show that the *same* conclusion R follows no matter which of P or Q happens. Then you can conclude that R happens. For example, you know that it is going to either rain or snow. Now, if it rains you need an umbrella, and if it snows you need an umbrella. So, you can be sure that you're going to need an umbrella either way. The basic idea of Existential Elimination is the same as Disjunction Elimination. Remember that (∃x)Fx amounts to saying either Fa is true, or Fb is true, or …. In the sub-proof, we show that a desired conclusion R follows when Fa is true. If we could show that the very same conclusion follows when Fb is true, when Fc is true, and so on, we could conclude that R happens no matter what. But we can't have a separate sub-proof for every proper name, because there are infinitely many of them. In the case of Disjunction Elimination, there were only two of them involved, but here we have an indefinite number of instantiations. To solve this problem, we have introduced conditions 1-3. They make sure that our conclusion of R was not derived from any special features of "a". In other words, conditions 1-3 make sure that R would still have followed had we picked proper names other than "a".

Let's examine some examples that do violate at least one of conditions 1-3. Here is one:

1.	(∃y)(Ry . Ty)	A
2.	(Rb . Tb)	A (for ∃E)
3.	Rb	2, .E
∴	Rb	1, 2-3, ∃E

Incorrect application of ∃E:
"b" occurs in the conclusion

This proof fails to satisfy condition 1. The letter "b" occurs in the conclusion, in violation of our rule. Everything else is correct in this proof. Intuitively, the premise says that some R's are B's, but the conclusion says that b in particular is R. This is not a valid argument.

Now consider the proof below:

1.	(Rcc ⊃ P)	A
2.	(∃y)Ryy	A
3.	Rcc	A (for ∃E)
4.	P	1, 3, ⊃E
∴	P	2, 3-4, ∃E

Incorrect application of ∃E:
"c" occurs in an open assumption

Here condition 1 is not violated, because the conclusion is just P, which obviously doesn't have any proper names in it, let alone "c". But condition 2 is violated. The letter "c" occurs in an open assumption other than line 3, namely line 1. This makes "c" a special object. "c" cannot stand for any generic object, because we are assuming (on line 1) that c has a special feature, namely it is such that if it has the relation R to itself then P happens (say, "When Carol respects herself, we have big a problem"). Everything else is correct in this proof.

Finally, consider the following proof:

1.	((∃x)Rxx ⊃ P)	A
2.	(∀x)(∃y)Rxy	A
3.	(∃y)Rcy	2, ∀E
4.	Rcc	A (for ∃E)
5.	(∃x)Rxx	4, ∃I
6.	P	1, 5, ⊃E
∴	P	3, 4-6, ∃E

Incorrect application of ∃E:
"c" occurs in the original existential claim

This is a very interesting case. Condition 1 is not violated: "c" does not occur in P. Condition 2 is also not violated, because "c" does not occur in any open assumption other than line 4. Of course it does occur in line 3, but line 3 is just a regular line, not an assumption. It is condition 3 that is being violated in this proof. The letter "c" occurs in the *original existential claim* that the ∃E move is based on, i.e. (∃y)Rcy. This delegitimizes the proof. Everything else is correct. In particular, the move from line 2 to line 3 using ∀E is correct. So is the move from line 4 to line 5 using ∃I. What makes it wrong is the application of ∃E to a sub-proof whose assumption violates condition 3.

One final note. Always remember that our rules are only applicable to *main connectives*. For example, the main connective of premise 1 in the above proof is *not* the existential. It is the horseshoe. Therefore, you could not have used Existential Elimination on line 1. Likewise, the following proof is wrong:

1.	$((\forall x)Rxx \supset P)$	A	**Incorrect application of ∀E: ∀ is not the main connective**
∴	$(Raa \supset P)$	1, ∀E	

The mistake, as was said above, lies in the fact that Universal Elimination has been applied to a sentence whose main connective is not universal.

§8.E Homework

HW12

1- (0.5 pt.) Explain why we need some proper names to be arbitrary in proofs. Which rules require arbitrary proper names? Which condition in each of those rules secures arbitrariness?

2- (1 pt. each) Prove the following arguments. Be sure to check the conditions whenever there are any. Show that you have checked them next to the proof.

a. $(\forall x)Fx$, ∴ $(\exists z)Fz$

b. $(\forall x)(\forall y)Rxy$, ∴ $(\forall x)(\exists y)Rxy$

c. Sa, $(\forall y){\sim}Ry$, ∴ $(\exists x)(Sx \cdot {\sim}Rx)$

d. $(\forall x)(\forall y)Rxy$, $((\exists x)Rxb \supset Lab)$, ∴ Lab

e. $(\forall x)(Ax \supset Bx)$, $(\forall x)(Bx \supset {\sim}Cx)$, ∴ $(\forall x)(Ax \supset {\sim}Cx)$

3- (1.5 pts. each) Translate the following arguments into QL and prove them in derivations.

a. (UD: People) Everyone adores someone. Therefore, someone adores someone.

b. (UD: People) All Europeans are racist. Therefore, all Europeans are either racist or ignorant.

4- (0.5 pt. each) What exactly is wrong with each of the following proofs?

a.

1.	$(\forall x)Lxx$	A
2.	$(Laa \supset Bg)$	A
3.	Laa	1, ∀E
4.	Bg	2, ⊃E
∴	$(\forall y)By$	4, ∀I

b.

1.	$(\exists y)(\forall x)Rxy$	A
2.	$(\exists y)Ray$	1, $\forall$E
$\therefore$	$(\forall x)(\exists y)Rxy$	2, $\forall$I

c.

1.	$(\forall x)(\forall y)Rxy$	A
2.	$(\forall y)Ray$	1, $\forall$E
3.	Raa	2, $\forall$E
$\therefore$	$(\forall x)Rxa$	3, $\forall$I

HW13

1- (0.5 pt.) Explain the similarity between Existential Elimination and Disjunction Elimination in your own language.

2- (1 pt. each) Prove the following arguments. Be sure to check the conditions whenever there are any. Show that you have checked them next to the proof.

a. $(\exists x)Fx$, $\therefore$ $(\exists y)(Fy \vee Gy)$

b. $(\exists x)(Ax \cdot Bx)$, $\therefore$ $((\exists x)Ax \cdot (\exists x)Bx)$

c. $(\forall x)(Cx \supset Dx)$, $(\exists x)Cx$, $\therefore$ $(\exists y)(Cy \cdot Dy)$

d. $(\forall x)(\forall y)Dxy$, $(\exists x)(Dab \supset {\sim}Nx)$, $\therefore$ $(\exists x){\sim}Nx$

e. $(\forall x)(Fx \supset Gj)$, $(\exists x)Fx$, $\therefore$ Gj

3- (1.5 pts. each) Translate the following arguments into QL and prove them in derivations.

a. (UD: People) Someone adores everyone. Therefore, someone adores themselves.

b. (UD: Everything) Some birds can fly. Therefore, something is a bird.

4- (0.5 pt. each) What exactly is wrong with each of the following proofs?

a.

1.	(∃x)Lxx	A
2.	Laa	1, ∃E
∴	(∀y)Lyy	2, ∀E

b.

1.	(∃x)(Fx . Gx)	A
2.	│ (Fr . Gr)	A (for ∃E)
3.	│ Fr	2, .E
4.	Fr	1, 2-3, ∃E
∴	(∃x)Fx	4, ∃I

c.

1.	(∃x)Ax	A
2.	(∃x)Bx	A
3.	│ Ad	A (for ∃E)
4.	│ │ Bd	A (for ∃E)
5.	│ │ (Ad . Bd)	3, 4, .I
6.	│ │ (∃x)(Ax . Bx)	5, ∃I
7.	│ (∃x)(Ax . Bx)	1, 4-6, ∃E
∴	(∃x)(Ax . Bx)	1, 3-7, ∃E

Further Reading:

Teller, Paul (1989), *A Modern Formal Logic Primer*, available online, Volume II, Chapter 5

Merrie Bergmannn, James Moor, Jack Nelson (2009), *The Logic Book*, the McGraw-Hill Companies, Chapter 10

Chapter 9

Fallacies

§9.1 What Are Fallacies?

Throughout this course, we studied the principles of *good* argumentation. We learned how to make good, scientific conjectures, using the criteria of abductive logic (e.g. simplicity, testability, etc.). We also learned the features of a reliable generalization by examining the criteria that distinguish good and bad induction as well as various methods of criticizing and improving a given induction. In deductive logic, which we studied most extensively, we learned how to tell valid arguments from invalid ones (e.g. using truth tables), and how to prove ones that are valid. Most of this is about *good* arguments. Experience shows that people who know the basics of good arguments might still be prone to error in thinking. They might use or fall for bad reasoning, especially if it is put in deceptively persuasive terms. Thus, it is helpful to learn about *bad* arguments or, as we shall call them, fallacies, too. One can then strive to avoid using them as much as possible and train one's mind to resist their persuasive force.

Note that one does not learn about bad arguments simply by thinking about the opposite of good arguments. There is a lot more subtlety involved in the topic of fallacies. For one thing, there are infinitely many bad arguments that one could possibly make. The point of studying fallacies is to learn the ones that are most common and most persuasive.

Fallacies are often divided into two categories: formal and informal. However, it is hard to make the distinction between these two categories precise. The basic idea is that some arguments are fallacious by virtue of the fact that they follow an invalid *form*. Hence *formal fallacy*. With this definition, any invalid argument would be a formal fallacy. We have talked about some of these in passing in the foregoing chapters. A familiar example would be: "If the gas tank is empty, the car won't start. The car won't start. Therefore, the gas tank is empty". Other bad arguments are fallacious, not because they are invalid, but because they reflect bad debate tactics and/or mistaken or possibly malicious action in discussion. Hence *informal fallacy*. One simple example would be the following: Suppose your goal were to prove P, but you proposed an argument for Q – a claim different from P. Your argument for Q might even be valid, but you would not meet the expectations that you created in your interlocutor by proving Q (because he/she is waiting for an argument that proves P). You would thereby commit an (informal) fallacy. These examples should give you a rough idea about what distinguishes formal and informal fallacies. Let's study them in turn.

§9.2 Formal Fallacies

We have talked about *modus ponens* and *modus tollens* on several occasions. These two forms of argument are valid both intuitively and according to the rules of Propositional Logic.

$$\begin{array}{l} (P \supset Q) \\ P \\ \hline Q \end{array}$$

Modus ponens

$$\begin{array}{l} (P \supset Q) \\ \sim Q \\ \hline \sim P \end{array}$$

Modus tollens

Modus ponens says that since we know that P leads to Q and we know that P has happened, we are certain that Q happens. *Modus tollens*, on the other hand, says that since we know that P leads to Q and we know that Q has not happened, we are certain, in retrospect, that P could not have happened (if P had happened, Q would have followed, which is not the case according to the second premise). Both of these arguments are such that the premises guarantee the conclusion.

§9.2.1 Affirming the Consequent

You can create another curious argument with the symbols given in *modus ponens* by switching the second premise and the conclusion. What happens when we do that? The resulting argument reads:

$$\begin{array}{l} (P \supset Q) \\ Q \\ \hline P \end{array}$$

Is this a valid argument? The following example has the same form:

If you are a communist, you advocate government-run health care insurance.

You advocate government-run health care insurance.

Therefore, you are a communist.

This argument has the ability to convince a lot of people (and it has). However, this is an error. Just because P leads to Q, doesn't mean that P is the only thing that can lead to Q. Thus, simply because Q has happened, does *not* mean that P has. In the above example, just because communists advocate government health care does not mean that *only* communists have this view. Someone might advocate government health care for reasons quite different from communism. This is essentially the same as our good old gas tank example. It is true that if the gas tank is empty the car won't start. But if one day you realize that your car won't start, you cannot immediately conclude that you are out of gas. There might be other reasons.

This mistake is so common and so tricky that logicians have given it a name. It is called the fallacy of *affirming the consequent*. Luckily, the name is quite self-explanatory: The intention of the person who commits the fallacy is to conclude the *antecedent* by affirming the *consequent*. If you need definite proof that this is an invalid argument, you can use the truth table test to show that affirming the consequent is indeed a formal fallacy.

$(P \supset Q)$

Q

P

Affirming the consequent

(A fallacy)

A quick note about affirming the consequent: It might bother you that abductions sound an awful lot like affirming the consequent. For instance, you conclude that your friend is not home because he does not answer the door. You might characterize this reasoning as an argument like this: "If he is not home, he will not answer the door. He does not answer the door. Therefore, he is not home". A lot of philosophers have been bothered by this similarity. Some philosophers go so far as to say that abductions are bad arguments altogether because they are basically instances of the fallacy of affirming the consequent. This is a very old, very long, and very interesting puzzle in philosophy (especially philosophy of science). We cannot go into the debates surrounding this issue in this course. But one thing is clear: Abduction is a far more complicated process than a simple affirming the consequent. You do not conclude that your friend isn't home simply because that is one possible explanation for the consequent. You conclude that he isn't home because that is the *best* explanation for the consequent. Whether or not being the best explanation is enough to distinguish abduction from the fallacy of affirming the consequent is exactly the topic of the abovementioned old, long, and interesting debates.

§9.2.2 Denying the Antecedent

If you have an abstract mind, you might be wondering what happens if we do the same with *modus tollens*. That is to say, what do we get if we switch the second premise of a *modus tollens* with its conclusion? Here is the resulting form:

(P ⊃ Q)

~P

~Q

This is not a valid argument either. Consider this argument:

If you vote yes on Obama's jobs bill, you'll help save the economy.

You did not vote yes on Obama's jobs bill.

Therefore, you did not help save the economy.

As convincing as this argument may sound, its fallacious nature becomes evident once you think about it for a moment. It may very well be true that Obama's jobs plan would help the economy if it had passed. However, one might vote this bill down in order to vote for another plan that helps the economy even more. That they did not vote yes on Obama's plan is not enough information for concluding that they did not help the economy, because there might be other ways to do that. Here is a more intuitive example: "If you get an A+, you'll pass the class. You did not get an A+. Therefore, you did not pass the class". Obviously, one can pass the class with lower grades too. Getting an A+ is sufficient, but not necessary for passing.

This fallacy is called *denying the antecedent*. Again, the reason for the name is clear: The intention of the person committing the fallacy is to conclude the negation of the consequent by denying the antecedent. As the explanation above shows, this is an invalid argument. Skeptics can seek definite proof from the truth table test for validity.

$(P \supset Q)$

$\sim P$

$\sim Q$

Denying the antecedent

(A fallacy)

We will content ourselves with these two examples for formal fallacies. The rest of this chapter covers numerous types of *informal* fallacy.[1]

§9.3 Informal Fallacies

§9.3.1 Circularity or begging the question

[1] The following sections are modeled after Harry J. Gensler's Chapter on fallacies (*Introduction to Logic*, Chapter 15), mainly in order to match the exercises on LogiCola.

Let's start with what is definitely one of the most important fallacies. Remember that the purpose of an argument is to convince someone of something they do not believe. In order to argue for a desired conclusion, you need to make certain assumptions, which we call the premises of your argument. These assumptions better be agreed upon between the two parties (the speaker and the listener). If they are not agreed upon, they better be such that the speaker can independently *argue* for them and justify them. Now, in some arguments, the content of one or more of the premises is extremely close to the content of the conclusion. In fact, in some arguments, one of the premises contains a claim that, perhaps under closer scrutiny, is just what the conclusion says. In these cases, the argument defeats its purpose of convincing people of what they do not believe. Here is why. Suppose you are arguing for P. You present an argument where one of the premises claims something either extremely close to P or simply identical to P. In that case, you cannot convince someone of P by this argument, unless they already believe P. In other words, it seems that if the person were to accept your premise, he/she would already have accepted your conclusion, so there would be no need to make the argument. You are going around in a circle with the argument. Such arguments are called *circular* or *question-begging*.

An example would make this clearer. Suppose Sam, a believer, is trying to convince Sally, an atheist, that God exists. Sam makes this argument:

Bible says that God exists.

If Bible says something is true, it is true.

Therefore, God exists.

Sam's argument is valid and the first premise is clearly true. So Sally wonders what justifies the second premise. Sam responds by saying "Well, Bible is the word of God and it must be true, because God does not lie". This response makes Sam's argument *circular*. In order for Sally to accept the second premise of Sam's original argument, she has to believe that God exists. But that was exactly what the original argument was supposed to convince her of! How can the argument shown above convince a non-believer of God's existence, if they have to believe in God in order for the argument to be sound in the first place? Note that the problem with this argument is *not* that it deals with God's existence and that is a matter of personal opinion (there are other, perfectly fine arguments for the existence of God). The problem is that this particular argument is circular or, as philosophers would say, *begs the question*. Also, as said above, Sam's argument is valid, so he is not committing a formal fallacy; just the informal fallacy of begging the question.

The cause of Sam's fallacy seems to be that he is really caught up in his worldview, so he takes certain things for granted and forgets that other people might not believe those things. Thus, he uses one of those granted beliefs as a premise, not realizing that only people who already believe his desired conclusion can accept that premise. However, circularity is a complicated issue and even great philosophers are sometimes accused of begging the question – and sometimes rightfully so.

§9.3.2 Ambiguity or equivocation

Our next fallacy is ambiguity. As the name suggests, this happens when a term is ambiguous throughout the argument. The most interesting cases of ambiguity happen when a term switches meaning from one premise to another. Consider this example:

Justice is beautiful.

All beautiful things are such that they please human eyes or ears.

Justice is pleasant to human eyes or ears.

It seems that both premises are in some sense true and yet the conclusion is false. Justice is not pleasant to our eyes and ears – it is not physically sensible at all! If you blindly translate the argument into QL, it will turn out to be a valid argument. So how can the premises be true and the conclusion false? The trick is that "beautiful" switches meaning from the first premise to the second. In the first premise, "beautiful" means good, morally admirable, or some such. In the second premise, "beautiful" means physically attractive. These are two quite different senses of "beautiful" and cannot be interchanged. If we take the first meaning of "beautiful" and apply it to the second premise, it will become clearly false/dubious (morally admirable things do not please eyes or ears). Likewise, applying the second meaning of "beautiful" to the first premise makes it false (justice is not physically attractive). Therefore, the argument is *ambiguous* on the meaning of "beautiful" or, as philosophers would say, *equivocates on* "beautiful".

Of course this was a crude example. Sometimes, when a term has two or more very close meanings, people commit the fallacy of ambiguity without knowing it, and sometimes it is a challenging task to discover the ambiguity.

§9.3.3 Appeal to emotion

The problem with this fallacy is as self-evident as its name. Many people cannot tell the difference between an argument against a view and a simple condemnation of the view in emotionally loaded, derogatory terms. Consider what someone has said about the view called *determinism* (this is the view that says everything in nature, including our thoughts and actions, are caused by physical processes that precede them and is ultimately explainable by science.)

> Under this view, nothing is left to the person's free will to decide. We have no liberty in life. We are slaves of our brains. We are simply animals that do what the chemical reactions and the electrical pulses in their neurons tell them to do. This view degrades human beings to the level of insentient machines who cannot deliberately determine their own fate. What is worse, determinism leaves no room for morality and human sympathy.

The above paragraph can create a deep distaste for determinism in the reader. Yet there is not a single argument in this paragraph. The author has not provided any reason to doubt that determinism is true. All that the above sentences do is stir up a series of negative emotions that are somehow attached to some of the consequences of the view under consideration. Many terms in this paragraph are emotionally loaded, which reflects bad scholarship. If you have a cogent argument that can knock a view down, there is absolutely no reason to get emotional. You can simply present your convincing argument in purely descriptive terms. But more importantly, the fact that a view is, say, depressing or cruel has nothing to do with whether it is true or not. For all we know, the truth might be extremely depressing or cruel. Similarly, that a given claim is pleasant or describes a desirable state of affairs is no reason to think that the claim is true. Therefore, linking the consequences of a view with positive feelings is also an instance of the fallacy of appeal to emotion.

To argue for (or against) a view you need premises and conclusion. The conclusion must be something to the effect that the view in question is true (or false). Simply repeating what the view says in emotional terms is not an argument.

§9.3.4 Arguments that are beside the point, straw man arguments, red herrings

Debates can get long and confusing. Sometimes, in the middle of a debate, people lose track of what was originally the question and start arguing about something else. Say Sam and Sally, his wife, start a quarrel about how Sam always puts his family before Sally, for example by spending more time with them than Sally. After a little back and forth, Sam says: "I don't complain when you read your book for two hours every day". Sam's claim might be true, but it's beside the point. Even if he is right about that, and even if that is indeed something to complain about, it has nothing to do with the original problem, i.e. whether Sam thinks his family is more important that his wife. Sam's point might be that Sally is guilty of the same lack of involvement in her relationship, but even that is beside the point. Showing that the other person is guilty of the same thing does not absolve you. So Sam's claim is beside the point on all of these interpretations.

Sometimes people use this tactic intentionally, in order to distract their debate partner. In that case their argument is called a "red herring". (The name apparently comes from the way dog trainers used red herring fish to distract the dog from the smell it was supposed to follow.) Consider this debate:

- We, Republicans, stand for free market and are against government intervention in the economy.
- Oh, yeah? I guess prohibiting homosexuals from marrying whoever they want is not government intervention!

This is an example of a red herring. The first person (the Republican) is clearly talking about the economy. He/She claims that Republicans do not want the government to interfere with the economy. He/She has said nothing about interfering with other aspects of people's lives. In particular, he/she has not said that government intervention in social issues such as marriage is also bad. To throw the issue of gay marriage at them is merely a digression here. There is of course a legitimate question as to how being against government intervention in one case (i.e. the economy) can rationally go together with full support of government intervention in another case (i.e. gay marriage), but that is a different issue.

A special case of the fallacy of red herring or beside the point happens when one person mischaracterizes the other person's views. This is usually done by taking the other person's view to an extreme and arguing that such an extreme view is absurd. This is called a "straw man" argument. Here is an example:

- We, Democrats, stand for government programs that help the poor and the needy and protect consumers against corporate malpractice.
- So you are communists! You think that the government knows better than the majority of people and should make decisions for everyone.

Well, not quite! That is not what the first person said. Perhaps if you take the original statement and push it to an extreme you will end up with communism. But there is no indication in what the first person said that such an extreme is either intended or desirable. The second person is committing the straw man fallacy by mischaracterizing the view in question. There is of course a legitimate question as to where the boundaries of government regulation should be, and whether or not enacting such regulations opens the gate to all kinds of other top-down decision making, but that is a different issue.

To sum up, *beside the point* is the umbrella term for all instances where someone digresses from the original debate by stating something irrelevant. This action is called a *red herring* if it is meant to *distract* the partners in the debate. The same fallacy is called a *straw man* argument if it involves mischaracterization (usually in the form of exaggeration) of the view under consideration.

Note that one might present a valid and sound argument for a conclusion that is beside the point. Being beside the point is about the *conclusion*, not the argument proposed for that conclusion.

§9.3.5 Appeal to crowd

Humans are social animals and define themselves through integration with the society around them. For this (and perhaps other reasons) it is very difficult for people not to believe in something that everyone around them believes in. If you hear from a lot of people that something you think is true is not true, you inevitably start having doubts about your belief. However, from a logical point of view, the fact that everyone thinks something is true/false has absolutely nothing to do with it actually being true/false. There was a time when everyone in the world thought that the Sun turned around the Earth and everyone in the European Continent believed that the Earth was flat. Yet none of that was true – or so everyone thinks today.

The *fallacy of appeal to crowd* consists of presenting as *evidence* for the truth of P the (presumed) fact that everyone thinks P is true. A while ago, I was doing some research about the influence of corporate lobbies on legislative decisions in Washington, D.C. I talked to several people who believed that lobbies have a significant power on Capitol Hill. A lot of them could not present any real evidence for this claim. Instead, they got frustrated and said: “Ask anyone! Everyone knows that!” This is a clear example of appeal to crowd.

It is important to distinguish appeal to crowd from a similar, but much more legitimate abductive argument to the following effect: “Most (or all) people think P is true. Most of these people have had a fair amount of access to the pieces of evidence and argument relevant to the case. Therefore, there is probably some truth to P.”. This is an abduction in the sense that the truth of P seems to be a very good – perhaps the best – explanation for the observation that most or all people believe P. This will not be a good explanation, however, if the people have not examined all available pieces of evidence and argument relevant to P. Moreover, just like any other abduction, the conclusion is never 100% guaranteed by the premises. It is still *possible* that everyone believes P after carefully examining all the relevant evidence and arguments, and yet P is false. That is why one should be cautious about this conclusion.

§9.3.6 Opposition

A similar effect arises when people identify a group in their society as the enemy and define themselves as being different from them. It is very tempting to think that everything your enemy believes is false. Just as we said in the case of appeal to crowd, the fact that your enemy thinks P is the case has nothing to do with the truth or falsity of P (no matter how evil the enemy is).

The *fallacy of opposition* occurs when one argues that something is false because one’s opponents believe it. This – like many other fallacies – happens a lot in political discourse. On conservative talk shows you often hear that something this or that left-leaning politician said is very similar or the same as what Lenin, Stalin, or Karl Marx once said. Similarly, when you talk to progressives, you should expect fierce criticism if what you are saying resembles something a conservative politician or activist has claimed. It is important to remember that negative feelings towards the views of one’s opponents are no evidence that those views are mistaken.

§9.3.7 Genetic fallacy

Suppose you realize that your debate partner believes in P because of some purely psychological reasons. For example, it could be that he grew up in a family that firmly believe P and that’s why he believes it now. Or, on the flip side of that, it could be that when he was a little child, his dad forced ~P on him and he really hated it; so when he grew up, he found himself strongly opposed to ~P and became a dedicated advocate of P. Does any of that mean that P is false? Obviously not. Why someone believes a proposition has nothing to do with its truth value. Nevertheless, in a debate context, it is very tempting to speculate about the psychological origins of someone’s views in the attempt to refute those views.

The *genetic fallacy* is any argument to the effect that some proposition P is false because the belief in P has questionable psychological origins. This is one of the most interesting fallacies and it takes a lot of effort to avoid it. Here are some examples. Some economists argue that the most efficient economic system is a free market. Free market skeptics have a well-known line against these economists, which says that these economists are themselves very wealthy and are trying to protect themselves by forging evidence for *laissez faire* economy. Interestingly, people on the other side of the political spectrum, often make similar claims about climate scientists. They say that these scientists are falsely claiming that fossil fuels are dangerous to the environment because they gain from these claims politically. Such arguments make the logician uncomfortable. Ultimately, the *motivations* of the scientists do not matter. If you think their arguments are flawed, you need to look at the arguments themselves and find out what is wrong with them. If truth is on your side, it should be possible to pinpoint the bias in the evidence or the false claims in the arguments. For example, if you think that the economist or the climate scientist is lying for political reasons, you can simply look at their peer-reviewed journals and see where their arguments go wrong or where their evidence is unreliable. Nothing in this process hinges on the psychological origins or the interests of the argument-maker.

Sometimes the genetic fallacy is not personal, but historical. For example, someone might argue that the origins of philosophy go back to sophistry in ancient Greece and therefore philosophy is all sophistry.

But be careful. The persuasive force of the genetic fallacy is not completely baseless. Let us distinguish two things here. A simplified formulation of the genetic fallacy would be this:

You have psychological/political/etc. reasons for believing P.

Therefore, P is false.

As we said, this is a fallacy. But one can make a much better argument that is somewhat similar to the above argument:

You have psychological/political/etc. reasons for believing P.

Therefore, unless you have better arguments, you are not *justified* in your belief that P (even though P might still happen to be true).

When the psychological or political origins of someone's beliefs are exposed, we have some reason to doubt that they have good reasons for believing what they believe. Even though what they believe might happen to be true, they might not be *justified* in believing it. In the extreme cases where those psychological or political roots are the *only* things that cause the person to believe what they believe, we have pretty strong reasons for doubt. Still, this doubt is not about the *truth* of what they believe, but about whether they are *justified* in believing it. Here is an example. Suppose it can be shown that one's parents' religious affiliation is the strongest predictor of one's own religion. This can be taken as evidence that the main – perhaps the only – reason most people have for adhering to a particular faith is that they feel at home with it. Does this provide reason to doubt the existence of God or whatever other claims are made in that particular faith? No. For all we know, God might exist despite the fact that people believe in Him for purely psychological reasons. What we *do* learn from this, however, is that most people are not *justified* to believe in God until they find real arguments for this belief. Whether or not there are real arguments for the existence of God and whether these arguments are successful is a whole other story.

To summarize, the genetic fallacy happens when one confuses truth with justification, and argues that just because someone's beliefs can be explained psychologically, therefore they are false. Notice that the person who believes something for purely psychological reasons is not the one committing the genetic fallacy (they're being manipulated, but they're not committing any fallacy of argumentation). Rather, the fallacy is committed when someone draws on psychology to conclude that something is false.

§9.3.8 Appeal to Ignorance

It is one thing for us to be able to prove P; it is another for P to be true. There are many true propositions that we have not proven yet. There might even be some true proposition that humans can never prove. Likewise, it is one thing for us to be able to disprove P; it is another for P to be false.

The *fallacy of appeal to ignorance* occurs when one argues that P is false because no one has been able to prove P, or that P is true because no one has been able to disprove P. For example, it would be an example of this fallacy if one were to claim that aliens don't exist, because no one has ever proven aliens.

A comment similar to the one I made about the genetic fallacy is due. We should distinguish the following arguments:

No one has ever proven P.

Therefore, P is false.

No one has ever proven P.

Therefore, until better arguments for P are provided, we are not justified to believe P.

The former is the fallacy of appeal to ignorance. The latter is a much more reasonable argument. Once again, the key distinction is between truth and justification.

§9.3.9 *Post Hoc Ergo Propter Hoc*

This Latin phrase means "after this, therefore because of this". The point is that simply because B always happens after A doesn't mean that B is caused by A. The point can be generalized to all kinds of correlation that are not necessarily temporal. In modern times, people usually express this by saying "correlation is not causation".

The *fallacy of post hoc ergo propter hoc* is basically everything we studied in Chapter 3. So I am not going to elaborate on this fallacy here. But to give a very brief recap, we learned that before we can conclude that A causes B we must do some very careful examination of our samples, make sure that there is no bias built into them, that they are not too small to reflect the variation in the general population of objects, and finally that there is no uncontrolled factor, no common cause, and no backward causation. It is never possible to eliminate all biases, and all uncontrolled factors, etc., but we have to do our best. Multiple examples were provided in Chapter 3.

§9.3.10 Division-Composition

Something that is true of a whole might not be true of (all of) its parts. Conversely, something that is true of the parts might not be true of the whole. Consider this arguments:

America is the greatest nation on Earth.

Therefore, everything we have in America is the greatest.

This argument is invalid, because the premise might be true while the conclusion is false. It might be true that, on average, America is the greatest nation on Earth. That is to say, all things considered, it is the best place to live, for instance. This is perfectly compatible with the existence of *some* things that other nations do a better job at. As long as the pluses outweigh the minuses more than any other nation, the premise can be true. The conclusion, however, seems to claim that there are no minuses, which is not guaranteed by the premise. Another example would be:

This object is a chair.

Therefore, every part of this object is a chair.

This is obviously wrong. Some properties, such as the property of being a chair, emerge when parts are put together that did not exist in each part individually. These are examples of the *fallacy of division-composition*.

The fallacy can go in the other direction as well. That is, one might mistakenly conclude that the whole has a property because all of its parts do. Here is one example:

Every player in this football team is excellent.

Therefore, this is an excellent football team.

This is also an invalid argument. If the players do not play in harmony and coordination, the team might in fact be very awful. The properties of the individual parts do not necessarily carry over to the whole.

That is not to say that they never do. Sometimes the whole has some of the same properties as the parts. For example, if every part of a wall is red, the whole wall will be red. As an example of the other direction of the fallacy, it is possible to conclude that every part of a drink is pure water from the assumption that the whole drink is pure water. But the logician's point is that this *pattern* of reasoning is not *guaranteed*. It might work in certain cases, but not in others. When it doesn't, it is a fallacy.

§9.3.11 Appeal to Authority

One cannot research or go see everything for herself. There are many propositions that we believe, not because we have seen their proof personally, but because the experts on the subject say so. For example, I believe that the Earth is round. But I haven't seen the Earth from outside. Neither have I done any of the geometrical experiments that are used to prove this roundness on the surface of the Earth. I have this belief based on a proper appeal to authority: If all physicists and astronomers honestly and unanimously say that the Earth is round, then I believe them.

There are, however, improper appeals to authority as well. The *fallacy of appeal to authority* is committed when one or more of the following conditions occur.

1) The authority in question is not an expert on the subject at all, but an expert on a different subject or perhaps simply a celebrity. For example, the fact the President likes a certain kind of cereal does not mean that that is the best cereal.

2) The authority in question is an outlier and does not agree with the consensus of the rest of the expert community. For example, one climate scientist somewhere out there who believes that fossil fuels are OK and the environment is not being destroyed, or one economist somewhere out there who believes that communism is the best economic system.

3) The conclusion is overstated. For example, some people conclude that P *must* be true, or that P has been *proven*, or that anyone who doesn't believe in P is crazy, simply because all authorities on the subject think P is true. This is an inappropriate appeal to authority. The consensus of experts is almost never universal (there are always some disagreeing experts) and even if it were universal, it would not mean that P must be true or has been proven. There was a time when all experts unanimously believed that the Sun turns around the Earth. So experts can be wrong. The proper conclusion is that P is *probably* true, or that we should believe in P until evidence to the contrary is found.

To summarize, appeal to authority is allowed as long as the authority in question really is an expert on the subject, is not an outlier, and one does not overstate her conclusion.

§9.3.12 *Ad Hominem*

When we like the person who is talking, we tend to be more charitable and more trusting in our interpretation of what he/she says. As a result, we tend to believe his/her claims more easily than those of someone we don't like. This is a natural psychological reaction, which logical people should nevertheless try to avoid as much as possible. The fallacy explained in this section partially rests on this psychological effect

The *fallacy of ad hominem* occurs when a party in the debate attacks the personality, background, moral character, or credibility of his/her interlocutor, instead of attacking the view proposed by them. “Ad hominem” means *against the person* in Latin. Here is an example:

- Chastity and loyalty are the foundations of every family.
- Is that why you cheated on your wife?

The first person is making a general point about the values that he thinks are the foundations of family. Whether or not he himself has been able to live up to these standards is irrelevant. The second person is trying to ruin the credibility of his interlocutor by questioning his character or honesty. However, this personal attack fails to establish that the point the first person made was wrong. This is therefore an example of *ad hominem* or personal attack.

Here is another example:

- I think there are some things to be said in favor of President Bush’s record.
- Yeah, I remember you also defended Hitler in your interviews during World War II.

Again, the second person seems to be attempting to destroy her interlocutor’s reputation. But in fact, whether or not the speaker has defending Hitler on his record does not say much about the cogency of his arguments regarding Bush.

Less sophisticated examples of *ad hominem* happen when someone calls their debate partner “stupid”, “backward”, or “uneducated”. This type of *ad hominem* comes close to another fallacy called *appeal to force* (see below).

There might be more legitimate arguments that look similar to *ad hominem*. For example, if a person has a long record of holding blatantly false beliefs and/or obvious insensitivity to data and evidence, it is plausible to *expect* his other views to be misguided as well. This is a sort of induction about the person’s rationality and critical thinking skills. As long as we use the conclusion cautiously (“He/She is *probably* misguided about this one too”), this kind of reasoning has some legitimacy. Nevertheless, remember that such arguments are not well received in debate contexts. If you start listing the person’s other views and talk about how crazy they are, in order to conclude that his current view is also probably wrong, you will likely be accused of *ad hominem*. Try to find better arguments that are entirely non-personal and directly about the view under consideration.

§9.3.13 Appeal to Force

One commits *appeal to force* when instead of providing an argument, one threatens his/her interlocutor or tells them to "just shut up". Less extreme cases might include calling your debate partner "crazy" or telling them that they must first educate themselves before they talk (this fallacy sometimes goes hand-in-hand with *ad hominem*). If you have a good argument there is no reason to appeal to force. Authority figures like parents and teachers are perhaps more prone to this attitude when they cannot explain something to their children or students.

§9.3.14 Pro-Con

When evaluating something, whether it is a view, an action, a political party, or a policy, one needs to be patient and consider all pros (reasons in favor) and cons (reasons against). Once all pros and cons are listed, we can ask whether the pros outweigh the cons or vice versa. If we only list the positive aspects of something, it is easy to be lured into concluding that it is a good thing, because we are not actively thinking about the possible negative aspects. Similarly, if we only list the negative aspects and don't think about the potentially positive ones at the moment, we might feel that there are no positive aspects.

The *fallacy of pro-con* occurs when one lists only the pros or only the cons of something and omits the other side of the story. The tricky thing about this fallacy is that when one party in the debate continues to see the full half of the glass, the other party is likely to be dragged into seeing the empty half all the time to balance the argument. For example, when one party only talks about the benefits of capitalism and refuses to acknowledge that there might be some flaws in the system, the opponents are likely to starts listing all the evil that has ever happened under the capitalist system without appreciating the benefits of capitalism. This polarizes the debate, which is not constructive.

§9.3.15 Black and White Thinking

Our last four fallacies, including this one, are not really styles of reasoning, but more like attitudes that are not helpful to have in a debate context. So we cannot put these fallacies in the premise-conclusion form. The first one is *black and white thinking*.

People like to have definite answers for their questions. Ambiguity is unpleasant. For this (and perhaps other reasons) people often develop a picture in their minds that has a couple of opposing categories and quickly classify everything under either this or that category. Reality, however, is much more complex. Things often do not perfectly fall under either category.

The fallacy of *black and white thinking*, as the name suggests, occurs when one behaves as though everything has to belong to either of two extremes. For example, some people label everything that is not capitalism as communism or everything that is not communism as capitalism. In reality, though, there is a whole spectrum of possible and actual economic systems ranging from fully free markets to entirely government-owned economies. Most countries are somewhere in between the two extremes. Appreciating the fact that everything is not black and white – which is, by the way, easier said than done – can be very helpful; failing to appreciate this fact, on the other hand, can be very destructive in discussions.

§9.3.16 False Stereotype

We probably adhere to stereotypes for a lot of the same reasons as we tend to think in black and white. The world seems less scary when everything belongs to a known category. But stereotypes are often mistaken. It is important to try not to make generalizations about things too quickly. This fallacy also connects back to our discussion of inductions in Chapter 3. One should examine the sample and the method of induction very carefully before one can make a generalization about a group of people or objects.

False stereotypes are not always ethnic or racial (although that is one important example of it). We tend to make quick generalizations about a lot of things such as people in other neighborhoods, students of other majors, and food in a certain restaurant; for example, when we say things like: "All med students are nerds" or "All art students are really nice". Make sure you try it many times and under many different circumstances before you generalize.

§9.3.17 Complex Question

This is also strictly speaking not a fallacy, but an unhelpful debate strategy. A famous example of a complex question is: "Did you stop beating your wife?". If the person answers with "no", he has admitted that he beats his wife. But if he answers with "yes", he has also admitted that he used to beat his wife! If he never beat his wife at all, he cannot answer this question as is. A complex question such as this one presupposes something false and puts the person in a strange position where it seems that whether they answer in the positive or in the negative, they will have accepted that false presupposition.

§9.4 A Final Note

We have thus reached the end of our short journey in the world of logic and human reasoning. We briefly visited the lands of abductive logic – the study of theory-making – and inductive logic – the science of generalizing – and learned what they are and how they can be evaluated and improved. Then, we moved on to deductions and studied two main branches of deductive logic: Propositional and Quantificational Logic. In the former, we studied deductive arguments whose validity is due to the inter-sentential structures such as "and", "or" and "if-then". But in the latter, we focused on the sub-sentential structures and the kind of validity that depends on the relations between "all", "some", "none", and "some… not…" statements. We also examined various forms of fallacies and talked about how to avoid them.

There is a lot more to logic. Not only is each of the chapters you read merely the tip of the iceberg of its topic, but also there are many more topics and branches in logic that we neither studied nor mentioned. In the realm of abductions, we only examined 4 criteria of hypothesis evaluation and did not talk about several other criteria such as explanatory power, unification, intuitive plausibility, non-*ad hocness*, and so on. We also did not explore various philosophical questions about whether abduction is at all justified as a method for finding the truth. The status of observations, and in particular whether there is any clear line between observation and theory, is itself a very vexing problem in philosophy. As for inductions, we did not mention probability calculus, the Bayesian theory of induction, and the philosophical issues surrounding inductive reasoning in general. As an example of the latter, whether inductive reasoning is justified at all has been a major question in philosophy, at least ever since the 18^{th} century. In the arena of deductive logic, we did not mention, among many other things, Modal Logic (in which you can represent necessity and possibility), Deontic Logic (where you can represent commands and ethical statements), and many-valued logics (in which sentences are allowed to be "in between true and false" or neither true nor false). Nor did we exhaust the list of all important fallacies that have been studied by logicians and philosophers. If you are interested in further study of logic, you should refer to other textbooks. I referenced some other elementary textbooks in the *further reading* sections at the end of each chapter. There are also several advanced texts that I did not refer to but that you can easily find if you are interested.

§9.E Homework

HW14

1- (6 pts.) Complete R from LogiCola.

2- (1 pt.) Show that *Modus Ponens* and *Modus Tollens* are valid arguments while *Affirming the Consequent* and *Denying the Antecedent* are fallacies using a truth table for each of these forms of reasoning.

3- (0.5 pt. each) Write MP, MT, AC, or DA in front of each of these sentences to indicate that it is an instance of *Modus Ponens*, *Modus Tollens*, Affirming the Consequent, or Denying the Antecedent, respectively. Translate the sentences into PL or QL to help you find the answer. Remember that in logical translations, the apparent order of the English sentences does not matter. What is important is which sentence is the antecedent and which one is the consequent. It should also help to ask yourself: "What is the person's conclusion?".

a. Undergrads are lazy and slow. But I'm not an undergrad, which is why I'm not lazy and slow.

b. Philosophers are arrogant. That's why I think you can't become a philosopher because you don't have the arrogance in you.

c. I know you get angry when I don't listen to you. But I was listening this time, so you shouldn't be mad.

d. You spent all that money? You went to the mall again, didn't you? Because every time you go to the mall you waste all your savings.

e. He said I'll find him in the lobby. I'm going down stairs now, so hopefully I'll see him there.

f. Philosophers are always interested in these kinds of questions. Since he was asking that question, I figured he was a philosopher.

Further Reading:

Gensler, Harry J. (2002), *Introduction to Logic*, Routledge, Chapter 15

Copi, Irving M., Cohen, Carl, and McMahon, Kenneth (2011), *Introduction to Logic*, Pearson Education, Inc., Chapter 4