Introduction to Logic

INTRODUCTION TO LOGIC

JOHN SANTIAGO

College of DuPage Digital Press
Glen Ellyn

CONTENTS

INTRODUCTION

Target Audience

As a course of study in colleges and universities, logic is either a degree requirement or a strongly recommended class. While there should be no doubt that the study of logic benefits all students in higher education, one cannot overlook the fact that most of these students will never become professional philosophers, much less professional logicians. As such, this text is not meant to cater to discipline-centric goals: the reproduction of a scholarly minority and the preservation of intellectual traditions.

This text is meant for the majority of students found in logic classrooms, those whose intellectual pursuits lie far outside philosophy and the field of logic. Put differently, this book is for the 99% who will take a logic class and (should) expect to benefit from it.

The Approach of This Textbook

This textbook takes a student-centered approach. In this context, that means looking at the principles of logic as highly practical tools and training. The approach here is to treat logic as methodology that is actionable in everyday settings. This is logic for the street.

To the extent that some conventional techniques of modern logic are not conducive to everyday practice (e.g., the handwritten production of a 128-line truth table), we approach these in the context of the underlying value they do hold for everyday life. This text will cover those techniques. However, they are not covered here under the pretext that they are intrinsically valuable, and the text does not present them in order to ensure coverage of the discipline. The discipline can get along quite nicely on its own with the handful of scholars who cultivate it.

The coverage here is in service to students, most of whom really do need to learn this material (because their lives will be the better for it). Techniques such as these are framed in the context of enhancing student understanding of concepts, building mental focus, and improving rigor in communication. Broadly put, this approach speaks to the most basic value any student should find in taking a logic class: a more disciplined mind.

ABOUT THIS BOOK

This textbook was created for the PHILO 1120: Logic course at the College of DuPage with funding from the OER Faculty Support Grant Program.

Content for this textbook was adapted from the following open educational resources:

A Concise Introduction to Logic by Craig DeLancey is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License, except where otherwise noted.

Thinking Well : A Creative Commons Logic and Critical Thinking Textbook Edition 4.0 by Andrew Lavin is licensed under a Creative Commons Attribution 4.0 International License.

Fundamental Methods of Logic by Matthew Knachel is licensed under a Creative Commons Attribution 4.0 International License.

Bruce Thompson's Fallacy Page by Bruce Thompson. This material resides in the public domain.

Introduction to Logic and Critical Thinking by Matthew J. Van Cleave is licensed under a Creative Commons Attribution 4.0 International License.

Cover Image: "Trail, Foliage, Woods image" by Kaneori is free for use under the Pixabay Content License.

1.

WHAT IS LOGIC?

Basic Concepts

Logic: A Working Definition

Logic is the study of methods and means of evaluating arguments. Great. We have a definition in hand; now we can move on to the techniques.

Not so fast. Definitions like this are very convenient (you should memorize this one). They fit easily in the back pocket of your mind, and we can retrieve them on command at our leisure (or when the quiz question comes up). Yet in doing so we find it all too easy to repeat them mechanically and without understanding. Thus, they fail us when treated this way.

I will repeat: you *should* memorize this definition (as you should memorize all the definitions in this textbook). However, you should not *just* memorize it. A properly memorized definition is valuable because it works to organize additional concepts. A good working definition is like a paper or chapter outline: it strings together all the main components in one tidy piece of information. But make no mistake: if we have no command of those other concepts, then such a definition is empty, no matter how thoroughly memorized.

Logic is the study of methods and means of evaluating arguments. Let's break that down. Logic is...

> **The Study**: You will strive to learn—this is not a casual overview. This is not a movie-like experience in which you passively receive information. You will be the main mover in making this class worthwhile.
>
> **Methods and Means**: There are techniques used here. These techniques distinguish a logical mind from the less-than-reasonable minds so frequently encountered. These are the nuts and bolts of our discipline, which you should learn to strengthen your mind.
>
> **Evaluating**: There is a goal to logic—we judge. In the contemporary college climate, there is a bit of resistance to the very notion of judging people. Too bad. If you are not judging, you are not doing logic. *How well you do it*—that depends on your mastery of the methods and means.
>
> **Arguments**: There is a focus to logic—we look at arguments. Of course, everyone knows what an argument is...right?

Argument: A Working Definition

We might hope that everyone knows what an argument is when they see one. Experience has shown that this is likely a foolish hope. If you think this is obvious, if you think everyone gets it, and especially if you think you do not need to work at getting better at this…well, there is that fool's hope to cling to if you are so inclined.

Here lies the first really great task in this course of study: understanding what an argument is and how to identify one. First, a (very) rough definition:

An argument is a set of statements, some of which are intended to support another.

Let us try a quick example of what that might look like. Here's a set of statements:

1. College can be expensive.
2. Many students do not have the money to pay for expensive books.
3. If every class requires an expensive book, it is hard for students to take a full-time load.
4. Using free textbooks helps many students.

All by themselves, we don't really have an argument yet. We merely have a group of statements with very little structure. You could conceptualize it like this:

1. Statement
2. Statement
3. Statement
4. Statement

To the logician, there's not much of interest here. We don't have a sense of how this set is carved up. Yet if we frame it like this we may do better:

1. College can be expensive.
2. Many students do not have the money to pay for expensive books.
3. If every class requires an expensive book, it is hard for students to take a full-time load.

Therefore, using free textbooks helps many students.

That makes it easier to see the argument, and we can see how our current definition applies. We can see (a) the whole set, (b) the chunk of it we referred to as "some of which" now separated from the final part, (c) the remainder of the set, "another" statement. Most importantly, we can see that there is an important relationship between these two chunks. The first chunk (of three statements) was pitched with the intention of supporting the other chunk (the last statement). You can visualize an emerging structure now:

All of this is now easy to see thanks to the help of the word "therefore," which is represented above by the three-dot symbol: ∴.

So why did this change turn the original group of statements into an argument? We'll need to dive further into the definition of an argument to answer this.

Argument Components

In our first try at defining "argument" it may not seem like we got very far, but already we said a mouthful. Let's break this down. An argument is…

A Set: At the most fundamental level, an argument is just a (special) collection of things. You need that collection of things to make an argument. So when you see a single thing (as in "one statement") you should never refer to that as an argument. An argument is *the whole set*, and that set has special properties that a good logician will focus on. Indeed, a logician will pay more attention to the quality of the whole set than to the things most folks think are important.

Statements: Arguments are sets of statements—so we should realize that we are *not* talking about agendas or disagreements or fights. There is just a set of statements…with structure imposed upon it. Structure, as it turns out, is almost everything of importance.

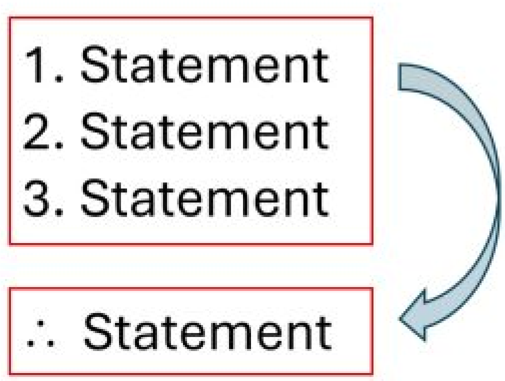
Some of Which: In the above group of statements, we find that some of them have been given a job. They are put forth by the author of the argument to get a job done. These statements work in the service of "another" (see below). They do *not* work in the service of *one another*. This block of statements often has its own structure, but at the most basic level we treat it like a gaggle of statements all pushing towards a common goal. Or at least, they *should be* all pushing in this direction.

Intended: Now we come to the heart and soul of an argument: there is a special intention to accomplish a very specific thing. The *author* of the argument has this intention. This is what brings forth structure into their collection of statements. As it turns out, this is the key to being able to identify an argument and distinguish it from any other group of statements. We (the *audience* who receives the argument) seek to understand the intention of the author. Once we see that they intend to use a group of statements as workers with a job that is aimed at another statement, we get our first glimpse of their argument. Of course, merely having an intention to do something does not mean one will succeed—our intentions often lead to failure or misinterpretation. This makes identifying an argument a challenge at times.

Support: The job of the first block of statements is to "support" another statement. That is their job; that is why they were brought into view by the author of the argument. Of course, having a job to do does not mean one succeeds in doing it. Having a job just means that a role exists for those with the job. The intention of the author is to put the first block to work: the job of "support" is to establish the truth of another statement. This sort of "support" is very different from "helping to explain another statement" or "giving an example of another statement" or "being about the same topic of another statement," etc. The meaning of "support" here is specific to truth: *on the basis of* the first block of statements, *the truth* of another *is established*. At least, that's the intention.

Another: The statement that is the target of the support has no job. That is, this statement is much like one who stands on the shoulders of others who try to hold them up. Ask yourself, who supports those on the ground? The answer is: nobody. Only this statement gets the intended support. The other statements in the argument are put forth unsupported. While this statement is much like a king who sits atop others who sweat and toil to support their glory, this statement is not really doing anything in the argument. This one is merely the final takeaway in the argument. This is the one claim the author was hoping you will believe to be true *because* they established its truth when they presented the other statements. At least, that's the author's intention.

You may find it useful to picture an argument as follows:



Now we can lay on some technical terms. But first...

Quick Sidenote: Technical terminology (i.e., jargon) is common in every field and discipline. We need to get comfortable adopting such terms, because (a) it helps us communicate clearly with others, (b) it helps us keep our own thoughts precise, and (c) due to (a) and (b) it helps us avoid errors and confusion. Besides, picking up jargon can be kind of fun sometimes.

Special Sidenote: In logic, some technical jargon is adopted from everyday language. That creates a high likelihood that the new student will get confused. They will conflate what they "think" a term means in logic with what they are previously familiar with hearing. *Do not do this*. When you hear a term you think you know, treat that feeling of familiarity with great suspicion. Chances are good that what you think something means has little to do with how a term is used in logic.

Back to our definition: we can insert some technical language.

> An argument is a set of statements, some of which (the **premises**) are intended to support another (the **conclusion**).

All we really did was add labels to the basic blocks that form the structure of an argument. We have not said anything more about these that we did not already know in the basic definition. For now, what matters is this:

There is nothing particularly special about a conclusion—it is nothing more than a statement, just like all the other statements in the argument.

There is nothing particularly special about a premise—for it too is a statement, much the same as every other statement in the argument.

The only thing that these labels do is help us keep track of the structure of the argument. And insofar as they do that, they do good work. For we need this technical language to help focus our mind while we are evaluating arguments.

Statements: A Working Definition

We know that arguments are just structured sets of statements. This definition does us little good if we do not properly understand what a "statement" is and how it differs from similar things.

In the simplest terms, a statement is simply a claim one makes about the world. You and I try to make claims about what's happening all the time. For example, we say things like:

- The game is on TV tonight.
- We are having spaghetti for dinner.
- I'm late for class!
- My car is faster than your car.
- This class is super cool.

Notice that for each of these statements, the attempt to make a claim about the world is *accountable to the* *world*. I mean, if the game is NOT on tonight, I got something very wrong in what I was trying to claim. I failed in an important way, because I was trying to say what the world is really like.

Statement: An assertion about some aspect of the world.

This is very different than other things we do with language. Consider the following:

- Will you watch the game tonight?
- Pass the spaghetti please.
- Why! Why! Why do I keep getting up late?
- Jump in the car.
- Take this class.

In each of these examples, we see that the speaker is not making claims about how things are in the world. Instead, they are (a) asking for someone else to make such a claim, (b) making requests for the world to be a certain way, or (c) simply expressing frustration. There is no effort to capture the way the world is or was.

Truth Value

The point we're making is that statements have a special quality about them. They are not simply English sentences ("sentence" is not a synonym for "statement"). Statements have a logical property we call "truth value." Since we are just starting our path to study logic, we'll keep this simple. For now, we will concern ourselves with only two truth values: true and false.

Statements have truth value. Questions, commands, and exclamations have no truth value. So, I can ask questions about their truth. Consider this:

"*True or false*: Some fleas are larger than elephants."

You might know the answer. You might not. In any case, at least you know that the question has a sensible answer. However, not every true/false question does, and there's a good reason for this.

Now consider how you would respond if I asked:

"*True or false*: Will you have spaghetti for dinner tonight?"

What if I asked this:

"*True or false*: Dogs?"

If you are confused by these last two "true/false" questions, good! That is pretty much the correct response. These things don't have a truth value. *Only statements have truth value*.

So now you know how to respond the next time someone asks: "True or false: Dogs?" The question makes no sense, because nouns have no truth value. Questions have no truth value. Requests have no truth value. Exclamations have no truth value. Only statements have truth value.

Let's get back to truth value itself. Some advanced systems of logic use more than two truth values, but for now the most intuitive starting point is to use just two: true and false. I say "intuitive" because most of us have a pretty good idea of what these mean (at least good enough to get us started).

TRUE: the truth value of a statement when it correctly describes the world

FALSE: the truth value of a statement when it does not correctly describe the world

That's going to work very well for us for a very long time. Besides, most people already walk around with this understanding in their mind.

As easily as this begins, there is no shortage of folks who claim to be confused about what is true and who gets to say that with any sort of authority. I suspect this is due to two confusions:

- Word games/misunderstanding what is meant by a word or a term
- Conflating *truth value* with *facts*

To the first confusion, we should make it a healthy practice of ours to ask for clarification when someone says something we think is wrong or does not make sense. All too often we get into unnecessary disputes when one person supposedly disagrees with another, only later to realize that they did not know what the other person was really saying. The imagined dispute over what is true or false ends up having nothing to do at all with truth value—it boils down to missing the reference of what was being said all along.

Reference

When we talk about "reference" in a statement, we are talking about what the statement is about. In short, the reference of a statement is its subject. In this sense, statements pick out a small bit of the world and try to describe it or its relation to other things. So, consider the following statements:

- The dog is sleeping.
- Aunt Sally makes great pecan pie.
- When you go to the store you always forget your shopping list.
- The housing market today is good for new homeowners.

In each statement, we should be able to find the thing that the statement picks out. This is what the statement refers to as it attempts to describe the world. The wakefulness of the cat or the hamster is not being described in the first statement—it is "the dog" which is pointed to and identified as the focal point of the description.

When folks disagree about what is being pointed to in a statement (e.g., is it "this dog Fido" or "that dog Spike" the speaker intended?), *and they don't notice they are talking about different subjects*, they often say they disagree on the truth of a statement (or worse, that truth is relative). In these cases, we should see that they do not disagree on truth at all. Rather, they disagree on which statement is being evaluated in the first place. A simple clarification would dispel the alleged dispute: I didn't say "Spike is sleeping," I said, "*Fido* is sleeping." Thus, the value of proper names should come into sharp focus here. Proper names are not always necessary, but crisp use of language makes the subject of our claims clear and helps everyone involved.

Facts of the World and Truth

To the second confusion, we should remember that only statements have truth value. *The world does not have truth value*—the world *be* whatever the world *be*, and those are the *facts*. If we distinguish between truth and facts, then we easily see that people can have legitimate disputes over how to properly make claims about the world, but they cannot dispute the facts of the world. If we look carefully, we see that so-called disputes about *the facts* are really just disputes about *our claims* of those facts. The facts are the very thing that help us judge between which claims are correct and which are not.

However much differing beliefs and claims may tear us apart, facts bind us together: they form a common point of reference we all share. We hold ourselves accountable to the facts in our efforts to make truthful claims.

> Facts are properties of the world.
>
> Truth value is a property of our statements.

Compound Statements

Before we leave statements, we will point out one more thing. Remember when we said that "sentence" is not a synonym for "statement"? This is important, as we will quickly see that many sentences contain multiple statements. Consider:

School can be challenging sometimes, but I am either up to the task or I will seek help when I need it.

If you go to the movie, you should not eat the popcorn nor pay for that overpriced drink.

Mike and Sally will sing in the play.

These are all complete single sentences. Each of these contains multiple statements. The speaker put smaller statements into certain relationships with other smaller statements to make a larger statement. When a speaker makes statements like these, we call them **compound statements**.

A compound statement is any statement composed of a logical relationship used on one or more simple statements.

Notice that to construct these compound statements, a speaker must do more than keep talking. The speaker must put smaller statements into logical relationships with other statements. When we speak out loud such relationships, we use key terms that express our intention. We need special words to set up these relationships in our statements.

Small words like "*and,*" "*if,*" and "*or*" do big work. These terms set up logical relationships, and in doing so, they establish the meaning of compound statements. That's huge! Unfortunately, people often fail to pay close attention to these important elements of compound statements. When they make this kind of mistake, they literally do not understand what the speaker is saying. Logicians try to avoid this error by focusing keenly on these tiny words. Later we will cover several techniques to pay very close attention to how these statements are constructed.

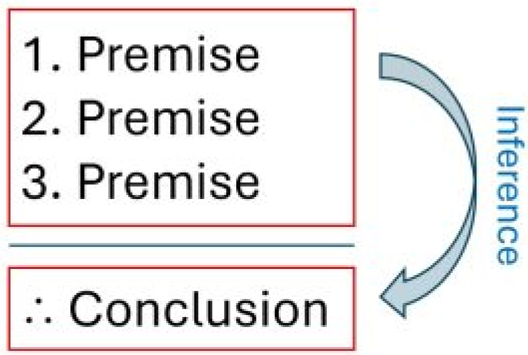
The Inference: A Working Definition

When we say that there is structure in a set of statements, we have seen that this comes from the intention of the author of the argument. The content of that intention (i.e., *what* is intended) is understood as "to support for the truth" of the conclusion.

Here's another way to look at the same intention: the author intends *to use* claims *to support* the truth of a statement.

The inference is *a leap of the mind* as it moves from considering one group of statements to asserting another statement.

You can visualize it this way:



Here's an example.

Let's say I tell you two things:

1. "Students who register for Mr. Marshall's class tend to get As."
2. "You *need* an A this semester."

Let that sit for a bit.

STOP! Before you go further, try reading the two statements again.

Q: Where does your mind go?

Chances are high that *on the basis of those two statements* your mind leapt to another statement:

"I should register for Mr. Marshall's class."

You might have doubts about the truth of the first two statements. But if you considered what the first two were stating, there's a good chance your mind leapt to this third statement anyway. That is, you inferred this *on the basis of* the other two statements.

NOTE: You did not "assume" you should register for that class—you INFERRED this *on the basis* of those statements. The idea that you should do so did not just pop into your mind *with no basis*.

Nor did this come to mind simply because you wanted it to be true or you hoped it would be true. Your mind did not make a leap of faith. We often misuse language in this way to say that what you now believe was simply "assumed," but that undersells your mind. Your mind *leapt* to this claim *from* the other two claims. An inference is not an assumption.

To be clear, whether or not your mind leaps from claims in reliable ways or leaps in wild flights of fancy is another matter altogether. Indeed, the study of logic is precisely the study of how to evaluate such leaps of the mind. The point is that our mind is active: it is not simply populated with beliefs, it is doing things with relationships between these beliefs.

Approaching Arguments

Up to now we have spent quite a bit of time trying to unpack the most basic nuts and bolts of what makes up an argument. Hopefully we have a better understanding of these components. Yet this does not guarantee that when we see an argument, we will be able to recognize that we are looking at one. This takes practice.

Our practice will be sped up by following a logician's order of inquiry. On a daily basis we encounter lots of things being said—we hear lots of statements. In the mix of all that, we hear a lot of arguments. While this may be so, we often hear them without realizing that we are hearing them. In short, we miss them. Yet this is better than the alternative: for it is far worse when we hear a fictitious argument, one that was *not made* but *we thought it was made*. See…this really does require practice.

A logician approaches a set of statements with a series of questions. First up: is there a special intention here? If not, move along…nothing much to see here. There are many times when we are just encountering a bunch of statements without any interesting structure.

If there is suspicion that we are running into a special intention, then the logician may first test their own uptake of the argument.

> If you say that you see an argument in a passage, then you should be able to identify what the conclusion is in the passage. If you cannot identify the conclusion, then you should be very hesitant to say you see an argument.

Often, we only suspect that we see an argument, because we only suspect that we know what the author's conclusion is in their statements. This is fine. Do not think that you have to automatically get it, like you were born with some kind of logician's magic. If you have confusion or uncertainty, it does not mean that you were not born with the logician's magic touch. With practice, we can grow.

When you suspect you know what the conclusion is in a set of statements, **test your candidate for the conclusion**. Ask yourself a question:

> Does it look like the author is trying to use the rest of the statements here to establish the truth of this candidate-conclusion?

Maybe you start to feel your stomach sink. You get your answer to this question, and you feel like "*no, they don't seem to be trying to do this*." In this case, your candidate-conclusion fails the test. That's okay. You may try another candidate.

Of course, you may also decide that there are no viable candidate-conclusions left and your suspicion about the passage was wrong: it is not an argument after all. Then again, you may find that the answer is "yes, the author *seems to want* to tell me the rest of the things that were said *in order* to establish the truth of this candidate-conclusion." If so, then you have a strong case for thinking this is indeed an argument.

> There is nothing automatic about this; there is nothing intuitive about it. We work the process through *careful use of questions* that help us identify the author's intention.

There is another way to approach this test of a candidate-conclusion. Ask yourself:

When I look at this candidate-conclusion, why should I believe it is true?

If the rest of what the author said seems to be trying to answer your question, then it gives evidence that this is indeed the conclusion, and this is indeed an argument. However, if the rest of what the author said does not seem to make the effort to answer, "Why should I believe this is true?" then you should back away from your suspicion.

Examples

Identify which of the following are arguments and which are not:

1. Strawberry ice cream combines the sweet, refreshing taste of real strawberries with a creamy texture, making it both delicious and slightly healthier than other flavors. Its vibrant color and nostalgic appeal add to its charm, making it a perfect treat for any occasion.
2. I think we should make a bird house. Making a bird house supports local wildlife, provides shelter for birds, and allows you to enjoy birdwatching right in your backyard. It's a fun, educational project that benefits both nature and your home environment.
3. Getting adequate sleep every night boosts your overall health, improves cognitive function, and enhances mood, ensuring you perform your best each day.
4. Riding a horse offers excellent exercise, improves balance and coordination, and provides a unique, enjoyable connection with nature and the animal. I'm glad to see Mary take up this hobby. She's making a good choice.
5. Turtles make good children's pets because they are low-maintenance, quiet, and can teach kids responsibility and patience through their care.
6. Look man, dogs are totally smarter than cats. Haven't you seen them? They understand more commands, perform complex tasks, and exhibit strong problem-solving skills due to their social nature and trainability. This is a no-brainer!
7. Watch Mike hit this ball. He's the best.
8. Recycling clothes reduces waste, conserves resources, and promotes sustainability, helping to protect the environment and reduce landfill usage.
9. They *say* voting is important for you. They *say* it empowers citizens to influence government decisions. They *say* it protects your rights. But all these politicians care about are your votes. None of these things is actually true.
10. ChatGPT can make mistakes due to limitations in understanding context, nuances, or up-to-date information, as well as the inherent challenges in natural language processing. You are just as likely to get a wrong answer from Chat as you are if you just took an educated guess. So, you should do your own work and make better-than-educated-guesses to answer questions.

Help and Challenges in Identifying Arguments

Illatives and Implicit Statements

Note that in the preceding examples, some terms may have helped you identify the intention of the author. We call these terms "illative" terms, which simply means they are flags that help identify an inference. Authors do not always use an illative term to signal their intention, but it sure helps when they do. For example,

So, (fill in the blank with a conclusion).

(fill in the blank with a conclusion), **because** (fill in the blank with a premise).

Thus, (fill in the blank with a conclusion).

Therefore, (fill in the blank with a conclusion).

Hence, (fill in the blank with a conclusion).

Consequently, (fill in the blank with a conclusion).

These terms make it pretty clear that the author wants you to believe in what they said precisely because they said other things they hope you accept. However, many times an author will use tone of voice, intonation, or other casual phrases to indicate their conclusions. Language is complex.

Indeed, things can get really tricky when an author intends to say something but does not say it out loud. That is, we often hear an author say something and know that they also meant to say something else along with it—they just didn't. These are implicit statements.

Implicit Statement: a statement that was not explicitly stated out loud or written down, yet can be charitably and reasonably included in an author's set of statements.

Identifying an implicit statement can be very difficult, and so it often gets us into trouble. We might be accused of putting words in the mouths of others.

"*I didn't say that!*"

"*Yeah, but it sounded like you were saying that…*"

Yeah. This is going to be a long night.

Logicians should be somewhat loath to claim someone is making an implicit statement. At the very least, you should be very careful and conservative in claiming that someone intended an implicit statement to be part of what they were saying.

The definition tells us we should "charitably and reasonably" be able to attribute the unspoken statement to the author. To be "charitable" we need to understand that the author may want to have included the statement but just couldn't or didn't have time, and so forth. So that rules out things like: "*You hate my mother!*" This is not likely something someone wants to say. That's just looking for a fight.

To be "reasonable" we need to consider the broad context of the statements. This is the really hard part. I might be able to *imagine* a scenario where the author *could have* intended to include a statement. This is actually very easy—and this is why **we are rightly accused** of putting words into someone's mouth. Do not do this. Imagination is not the same as reasonable judgment.

The fact that *you* can imagine an intention does not mean the *author* intended it. You are accountable to the author's words. If the words (or the actual context in which they are delivered) warrant claiming that there is enough context supplied, then we can cautiously include the implicit statement. However, if the words (or actual context) do not supply that context, our imagination is not a suitable substitute.

Consider two of our previous examples:

> "Getting adequate sleep every night boosts your overall health, improves cognitive function, and enhances mood, ensuring you perform your best each day."
>
> "Recycling clothes reduces waste, conserves resources, and promotes sustainability, helping to protect the environment and reduce landfill usage."

Neither one of these are arguments. However, many folks may think "*hey, I can see how someone might argue for this...*" In proceeding this way, we are using our imagination to override the author's actual words.

In the first, we might imagine a situation in which someone would say this in order to support a conclusion (something like: "You should get adequate sleep"). However, look at the passage again. Is there anything provided that tells us clearly that the author wants us to believe it is true that we should get adequate sleep? No. The author is just telling us the benefits of adequate sleep. This would likely appear in an argument with that conclusion, but nothing in the passage itself suggests that such an argument is being made. This is just a statement about the benefits of adequate sleep.

In the second, we might see even more clearly that this is just a statement about the benefits of recycling clothes. We can easily imagine a scenario in which a friend is trying to convince us it is true that "We should recycle our clothes." However, that imagined scenario is not what is directly in front of us. The words directly in front of us could just as easily be nothing more than an explanation of benefits. As audience members, *we have to push this into an intention that we fabricated*. The author's words alone don't give us enough evidence to make this judgment about *their* intention.

> **TIP:** When you find yourself pushing to create a backstory that justifies your claim to an author's intention, you are likely pushing too hard to create an intention that does not exist.

Consider another example from above:

> "They *say* voting is important for you. They *say* it empowers citizens to influence government

> decisions. They *say* it protects your rights. But all these politicians care about are your votes. None of these things is actually true."

Here we have an argument, and we have a good bead on the conclusion: "All these politicians care about are your votes." However, to really understand what the author intends, we will find it helpful to understand an important implicit statement in the passage. We should pick up on something like:

> "Politicians are lying to you."

We might even think this is likely included in the author's argument:

> "Politicians know these things are not true."

Unlike the other examples about sleep and recycling clothes, we do have something in front of us to point to when we say the author likely intends to include these in their statements. The repeated and italicized phrase "they *say*…" strongly suggests that the author wants us to understand that there is something suspect about what these people say. This amounts to an important premise in their argument. We are not using our imagination here to concoct some story about the author's intentions. We are looking directly at what they say and how they say it.

Given that the examples in this textbook are always provided out of context, you need to be able to point to specific aspects of the black-and-white text to make a claim for an implicit statement. However, in real life you often have the benefit of a live environment as well as the overall context of an extended conversation to provide additional support. The moral remains the same:

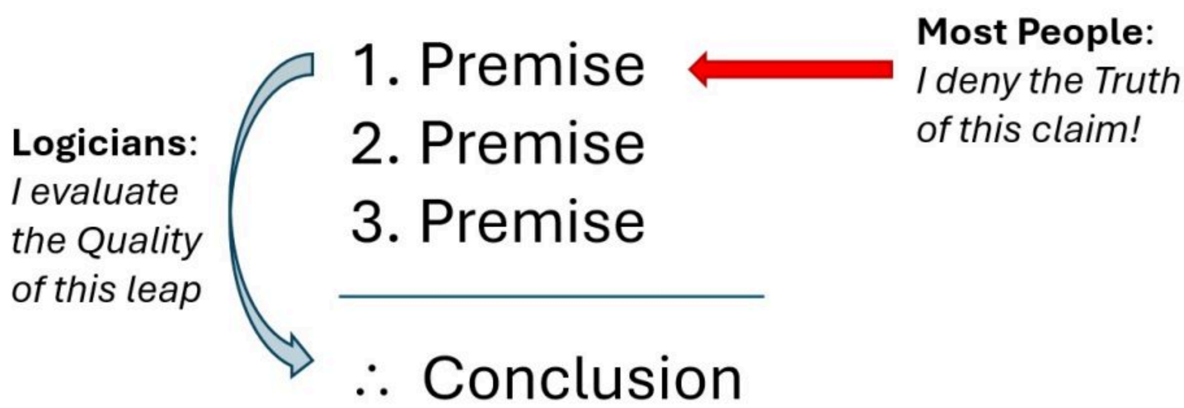
> Be very careful and cautious in claiming that someone intended an implicit statement, and be on guard against putting words in people's mouths.

Logical Properties of Arguments

Validity: A Working Definition

Once we know that we are faced with an argument, we can begin our evaluation of it. Note that when faced with an argument, most people will not evaluate the argument. Most people will *simply evaluate the individual statements* that make up the argument. Most people are not well trained.

Logicians do not judge individual statements first: they look to *the relationship between* the premises and the conclusion. They do not look at the statements as much as they look at the spaces between the statements (more importantly, the space between one chunk of the set of statements, the premises, and the other chunk, the conclusion).



The first concern to the logician is a key property of arguments: formal validity. This is a property of arguments—this is not a property of statements. So it makes no sense to talk of "a valid statement" nor "a valid point" made in an argument. Only an argument as a whole has this property.

The concept of formal validity is often misunderstood in large part because *it means less* than we think it means. So let's go slow.

An argument is valid if and only if **it is not** possible for all the premises to be true and the conclusion false.

That's not pretty. This account is awkward and lacks a graceful flow, but it works very well to help us avoid confusion. Of course, there are other ways to define validity correctly. However, these ways often mislead students into thinking validity is a robust concept…one familiar to their everyday preconceptions about validity. That is why we will not use these other definitions.

Let's try a simple example:

1. Mike and Larry will go to the football game.
2. Jane and Shauna will go to the football game.

So, Mike and Shauna will go to the football game.

Ask yourself: Are you sure that Mike and Larry will go to the football game? You probably don't want to bet the house on that claim. After all, *that single claim* could be false. So you probably don't want to bet that the conclusion is definitely true. However, *this is not what is at stake* when we ask if this argument is valid.

When we ask if this argument is valid, we just want to know if it is possible that you could say both premises are true *and then* deny that the conclusion is true. So…can you deny that the conclusion is true if both premises are true?

A: No (and if you think you can deny this, read the example again).

The punchline here is that we can see how we must accept the truth of the conclusion if both of those premises are true. So being able to give a valid argument is generally a good thing.

Going back to our ugly definition, the first thing we should notice about this definition is that it is framed in terms of a negation.

An argument is valid if *it is not*…

So, in this account the concept of formal validity is a kind of denial. When an argument is formally valid, we know for sure that something is NOT the case. Be very careful with how you think of things that are denied. Ask yourself:

Q: When you know that something is not the case, *what else* do you know?

For example, remember when you were younger, and you had to check in with your mom to let her know where you were hanging out? You told her the truth! (You would never lie to your mother…) You truthfully told her:

No Mom, I'm NOT at little Timmy's house. (She never liked Timmy anyhow.)

As a good and honest child, you spoke truthfully. Then again, in possessing this truth, what did Mom really know? Did Mom know where you *actually were* or just where you definitely could not be found?

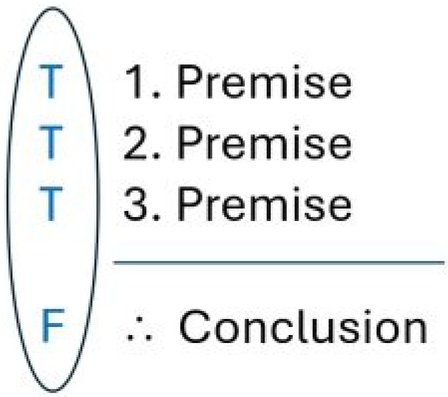
A: Momma knew very little about your actual whereabouts.

(…and you probably knew this, which is why you told her the truth *this* way…)

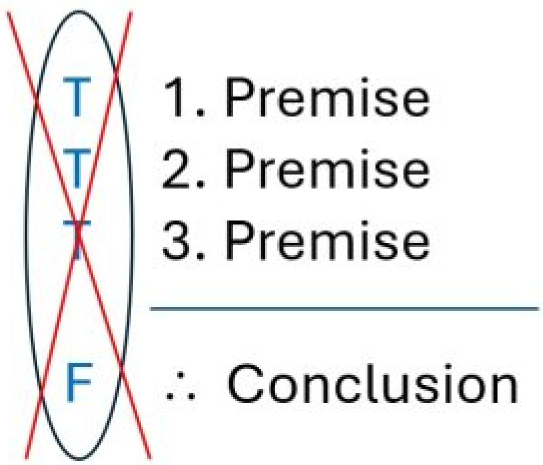
The same is true of formal validity. When an argument is valid, you know that something is definitely not the case, but you do not know what definitely *is the case*. You know very little about the argument. So don't get your hopes up for this concept; it is a thin concept with very little substance.

We cannot say a valid argument has a true conclusion, we cannot say a valid argument has true premises, we cannot deny that the statements are all false, we cannot insist on much of anything.

What we can say is that there is a kind of guarantee in a valid argument: the guarantee is that one thing in particular is not possible. Put differently, there is *a specific combination of truth values* that will never appear. Namely: all true premises and a false conclusion.



For a valid argument, you will NEVER see that happen. So you can visualize it like this:



Ask yourself: If I know that this combination of truth values will never appear, what combination of truth values does appear?

Like Mom, you cannot say. You do not know much of anything about the truth values in a valid argument. All you can say with certainty is that "T, T, T / therefore F" will never be the case. Still, this is valuable information. After all, as we saw above, *if it turns out* that the premises are indeed all true, then you know something important about the truth value of the conclusion: it *must also be* true. So, we can now safely say that another way to define validity is as follows:

Validity (alt. definition): An argument is valid when the truth of the premises guarantees the truth of the conclusion.

In saying this, we are not saying that the premises are actually true; we are simply saying that IF they are all true, this will lock in the truth of the conclusion.[1]

Identifying Validity

One of the easier ways to test for validity is to resist the temptation to attack individual statements in the argument. We want to know if something is not possible. Specifically, we want to know if the union of two things is not possible. We want to know if we can deny the possibility that these two things coexist:

(a) all the premises are true

AND

(b) the conclusion is false

Is this union of (a) **and** (b) impossible?

We do not (yet) want to know what is actually the case, just what is not possible. The simple way to test for this is to accept half of what is denied and see if we can make the other half feasible. This is what it looks like:

1. Accept the (a) part: treat all the premises as if they are true
2. Now see if you must accept the (b) part (while still holding on to the (a) part)
3. Ask yourself: If one really believed all of (a), is there any way to avoid accepting (b)?

This test reveals if you must accept the truth of the conclusion under the truth of the premises.

Invalidity: A Working Definition

Validity is an all or nothing thing. If an argument is not valid, then it is invalid. But what does that mean?

> An argument is invalid if and only if **it is** possible for all the premises to be true and the conclusion false.

Notice first that the definition of invalidity is almost the exact same as that of validity. This is because they are

1. The cautionary tale here is to remember that this is simply a formal constraint on the argument. There is no substantive relevancy requirement, as though it has to "make sense" that the conclusion follows from the premises. This is just a bare constraint on a combination of truth values, regardless of what substantive relationship exists between each premise or between the premises and the conclusion.

sister-concepts (two sides of the same coin). All we are inquiring into is whether or not a certain possibility exists—*the same possibility* is highlighted in each definition. Thus, our definitions are worded in almost exactly the same way. The only difference is whether *it is* or *is not* possible for that specific combination of truth values to happen.

Logical Categories of Argument Types

We have seen that once logicians know they have an argument in front of them, they approach it with a basic question. Is something possible or not? Once they have answered that, they know if the argument is formally valid or invalid. However, the logician's inquiry does not stop here.

Logicians systematically continue their inquiry based on the answers given to previous questions. We will do well to try to discipline our minds in this way. Logicians do not rush forward agreeing or disagreeing with the truth of the conclusion. After all, this is something anybody can do and shows no real skill at evaluating arguments. Rather, the logician will carefully look into the relationships that help us understand the quality of reasoning in the argument.

Soundness

Say we discover this about a given argument: No, it is not possible for all the premises to be true and the conclusion false. Now what question comes to mind? You are not alone if now you start to wonder: Well, *are* all the premises *in fact* true?

The answer to this question is *now* important to us. After all:

- If they **are** all true, **we have to accept** the truth of the conclusion.
- If they **are not** all true (which means at least one of them is false), then **we are not forced to accept** the truth of the conclusion.

Put differently, we see that there is a lot on the line now in whether or not all those premises are true. This is the logical force of a valid argument. Not surprisingly, this is also why many folks are so keen to poke holes in the truth of a premise. For if they can find fault in just one of the premises, they take all the wind out of the logical force behind the argument.

Logicians have special names for arguments that have these qualities. So if we already know an argument is valid, then we can say of the two options above that an argument is:

Sound: when it is valid *and* all the premises are true

Unsound: when it is valid *and* at least one of the premises is false

All things being equal, we regard the Sound Argument as our gold standard for quality in an argument. These are very powerful arguments that force all rational people to accept the truth of their conclusions. They are also very rare and hard to produce.

Strength and Weakness

Most of the time when the logician asks the first question of possibility, the answer reveals that an argument is formally invalid. This is not a death sentence for the argument. This does not mean that the argument is "bad" nor that we might not seriously entertain it. Indeed, the logician continues to consider and evaluate it. We might still have a pretty good argument in front of us.

So, a new question comes to mind when we know the following: Yes, it is possible for all the premises to be true and the conclusion false. The question is no longer of possibility but of probability.

Q: How probable is it that the conclusion follows from the truth of the premises?

Notice, we have already accepted that *it is possible* for the conclusion to be false (even if all the premises are true). But *how likely* is that? If it is not likely it will be false, then the conclusion probably follows from the truth of the premises. Let's face it, that's not bad. Absolute certainty is tough to come by these days. Having a high degree of probability is not as good, but it would be nearly as nice.

Arguments that are invalid are met with the question of probability. The result of our question leads to the following classification of invalid arguments:

Strong: an invalid argument in which the conclusion *probably follows* from the truth of the premises

Weak: an invalid argument in which the conclusion *does not probably follow* from the truth of the premises

To understand "probably" and "not probably," we can consider the percentage of likelihood. Anything that is 50% likely or worse is "not probable." Anything above 50% (even just a little, like 50.2%) can be said to be probable.

In practice, we often do not get firm percentage figures in an argument. More often than not, people will present arguments with indicators of probability. For example,

1. Most people who exercise daily live longer.
2. Mike exercises daily.

Therefore, Mike will live longer.

The term "most" is not as precise as 98%, however, we get ample sense of the probability that the conclusion follows from the truth of the premises. That is, *assuming* that the premises are true, we can say that the conclusion probably follows. We don't know that the premises are true simply because we know that the argument is strong.

We should note something similar and something different here from our discussion of validity/invalidity. The similarity is that in stating any of these, we do not also state that we know the truth value of any of the premises. The difference is that unlike validity/invalidity (which is an all-or-nothing property), strength and weakness in an argument admit of degrees. Like physical strength and weakness, the strength and weakness of an invalid argument vary a great deal. You might be strong but admit that there are folks stronger than you. You might be weak but admit that you are not the weakest kid on the block. All forms of strength and weakness come in degrees.

Cogency

If we know that an argument is strong, we have some assurance in the logical force of the relationship between premises and conclusion. We are not 100% confident, but we are well assured…that is, if the premises are all true. Now a familiar question should come back into view.

Q: Are all the premises true?

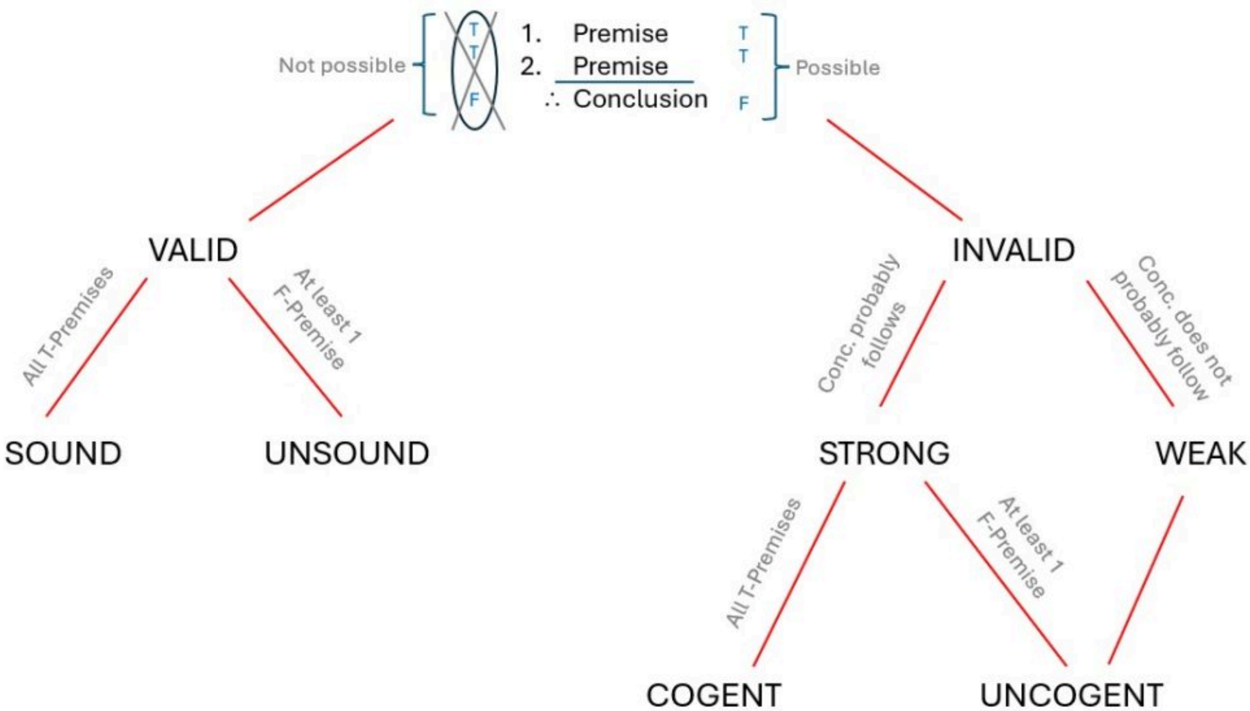
If the answer is yes, then we know we should probably accept the truth of the conclusion. In such a case the argument retains quite a bit of logical force. So, we classify these as such:

Cogent: an invalid argument in which the argument is strong and all the premises are true.

Uncogent: an invalid argument in which the argument is strong and at least one premise is false.

Similar to an unsound argument, when we learn that an argument is uncogent, it has lost much of the assurance that we were looking for in a quality argument. Our confidence that the conclusion is true is undermined, and the argument no longer appeals to our rational nature.

Additionally, we will regard all weak arguments as uncogent. This is in large part because they also lack all logical force regardless of the truth of their premises (because we cannot say that the conclusion follows even if all the premises are true). Weak arguments are pretty much worthless.



Arguing Well (or, Backing Out of an Argument without Backing

Down)

Let's go back to the beginning. If we want to mount a quality argument, we are probably going to try to make a sound argument. Yet we know that this is difficult to do. Often someone points out that we cannot really maintain the truth of all of our premises—they "poke holes" in one or more of our premises. Having an unsound argument is less than ideal, because we know that the audience does not have to accept our conclusion.

So do we give up? No! We fight back. This is an example of how it usually goes:

> Jeanine: "Listen Bob, you should register for Mr. Marshall's history class. Everybody gets an A! You take his class. You get an A."
>
> Bob: "Hold up. What about our buddy David? David took Mr. Marshall's class, and he didn't get an A."

Pause: Jeanine's argument is a bit more complicated in structure than it first appears. We'll look at that later. For now, here's the format of Jeanine's argument that concerns us.

1. Everybody who takes Mr. Marshall's history class gets an A.
2. You take Mr. Marshall's history class.

> Therefore, you will get an A in his class.

Bob has picked up on this argument and found an error in the first premise. Bob found a counterexample to refute Jeanine's key premise. Will Jeanine give up? Not likely.

> Jeanine: "Okay, fair enough, but we both know David's not the most serious student. Almost everyone who takes Mr. Marshall's class gets an A. Take the class. Boom! You'll get an A!"
>
> Bob: "Well, what about Hank? Hank took Mr. Marshall's class and he didn't get an A either."

Pause: Notice that Jeanine backed out of the attempt to construct a valid argument. Her strategy shifted and now she is intentionally constructing an invalid argument. However, Jeanine wants to make it as strong as possible.

Unfortunately, Bob has again found an error in Jeanine's key premise. Bob is chipping away at Jeanine's ability to maintain that all the premises are true. Does Jeanine give up? Again, not likely.

> Jeanine: "Okay, okay. We don't hang out with the sharpest crowd. But that's not you, Bob. Most people who take Mr. Marshall's class pass with an A. Take the class. You'll get an A!"
>
> Bob: "You're right. I'll probably get an A."

Pause: Notice that Jeanine softened the key premise from "almost everyone" to simply "most." If Bob kept pushing with counterexamples, this may have been softened even more (e.g., "better than half…"). However, Jeanine also squeezed Bob into the group of "most" by insisting that he is not like those others who fall in the class of Bob's counterexamples. This left Bob with good reason to accept the truth of the conclusion. Bob is not guaranteed to get an A, but if Jeanine is right about what is claimed in the premises, Bob will likely get an A.

Side Note on Inductive vs Deductive Arguments

Arguments that are invalid are evaluated using inductive standards of reasoning. We will look more closely at inductive arguments in later chapters.

Catalog of Argument Types

Valid: an argument is valid if and only if it is not possible for all premises to be true and the conclusion false

Invalid: an argument is invalid if and only if it is possible for all premises to be true and the conclusion false

Sound: a valid argument whose premises are in fact all true

Unsound: a valid argument with at least one false premise

Strong: an invalid argument whose conclusion probably follows from the truth of the premises

Weak: an invalid argument whose conclusion does not probably follow from the truth of the premises

Cogent: a strong argument whose premises are in fact all true

Uncogent: either a strong argument with at least one false premise, or any weak argument

Logical Properties of Related Entities

Now that we have seen how arguments with different qualities can be classified, we can look at other logical properties that interest logicians.

Statements

Statements can be classified according to important possibilities tied to their truth value. Three classifications are worth noting: logical truth, logical falsehood, and logical contingency.

We'll start with the first of these.

Logical Truth

Logical Truth (Logical Tautology): a statement is logically true if and only if it is not possible for it to be false.

Since we only have two truth values (true and false), if it is not possible for a statement to be false, then we can say it is always true. Notice how different this is from saying that a statement is merely true. "The student is reading" is a true statement. However, I could not say that it is not possible for that statement to be false. Clearly, that statement can be false; it is merely true right now (assuming our reader is a student). So a *logically* true statement is very special. Here's a cheap and easy example:

"The lights are on, or they are not on."

Clearly, that statement will be true no matter what the state of the lights is right now (or ever).

Logical Falsehood

Chances are that some of you reading can suss out the definition of a logically false statement. Give it a try before reading further…

Logical Falsehood (Self-Contradiction): a statement is logically false if and only if it is not possible for it to be true.

As before, since there are only two truth values, such statements are always false. Again, such statements are pretty rare and unusual. Another cheap and easy example:

"The lights are on, and they are not on."

Again, this example makes it easy to see that no matter what is going on with the lights, this statement will incorrectly describe them. There are far more interesting and complicated logically false and logically true statements. Unlike these examples, many are hard to identify at first glance. Later we will develop tools to evaluate statements to determine if they have these qualities.

Logical Contingency

Most statements that we encounter are not like logically true and logically false statements. These are logically contingent statements.

Logically Contingent (Logically Indeterminate): a statement is logically contingent if and only if it is possible for it to be true under some conditions and false under other conditions.

These are what we expect of statements. Sometimes it is true that the student is reading and sometimes it is false that the student is reading. Different conditions yield different truth values. This is part of where they get the names "contingent" and "indeterminate." The truth value of such statements is "contingent" (i.e., dependent on the state of the world). We can also say that unlike logically true and logically false statements whose truth value is fully determined by their own nature, the truth value of logically contingent statements is *not-yet-determined* by their own form.

Sets of Statements

Often logicians find it useful to understand the logical properties of a set of statements that is not an argument (or is not treated as such). Sometimes you just want to understand a group of statements as simply a group (without a specific structure).

Logical Equivalence

The smallest such group is a pair of statements. We sometimes find that two statements seem to track one another in their truth value. We call these statements logically equivalent.

Logical Equivalence: a pair of statements is logically equivalent if and only if it is not possible for one to have a different truth value than the other under the same conditions.

In an equivalent pair, when one statement is true, the other one is true. The same holds for when one is false—the other will be false. They never have *different* truth values.

This is not to say that one (and the other) must have this or that specific truth value. So do not make the mistake of thinking only contingent statements can be logically equivalent. After all, two logically true statements will always have the same truth value. Put differently, it is not possible for two logically true statements to have different truth values under the same conditions (because under all conditions they have the same truth value). So any logically true statement is logically equivalent to every other logically true statement. The same holds for logically false statements. While this is easy to see, when it comes to two logically contingent statements, figuring out if they are logically equivalent can be more challenging. For example, consider these two statements:

The door is open.

It is not the case that the door is not open.

Those statements are probably not too hard to identify as logically equivalent. In a sense, the two negations of the second statement cancel each other out and leave us with the same meaning as the first statement. Now consider these two statements:

If the dog is asleep, the cat will come inside.

Either the dog is not asleep, or the cat will come inside.

These two statements are also logically equivalent. A little bit of thought will reveal this, but it is not as obvious as the first example. Later we will develop tools to help us identify logically equivalent statements easily, even when they are quite complex.

Logical Consistency

When we look at larger sets of statements, we often want to answer an important question:

Q: Is it possible that all those statements can be true?

We might ask this question even when we are not looking at an argument. Then again, we might ask this question when we are looking at only part of an argument (for example, when we are looking only at the premises). This is a question of consistency.

Logical Consistency: a set of statements is logically consistent if and only if **it is** possible for all members of the set to be true under the same conditions.

The main thing to notice in this definition is that we are asking if a specific truth value is possible for all the statements—we are *not asking* if the truth value is merely "the same" for all the statements. We want to know specifically if they can *all be true* under the same conditions.

Logical Inconsistency

The savvy student may be able to suss out the definition of logical inconsistency. Give it a try.

You may notice that this is the second of our sister concepts. We previously noted that validity and invalidity were like two sides of the same coin. Thus, their definitions were almost exactly the same. The one thing that is at stake there was whether something is or is not possible. The same holds here.

Logical Inconsistency: a set of statements is logically inconsistent if and only if **it is not** possible for all members of the set to be true under the same conditions.

Notice two important things:

1. The definition changes and hangs on just one thing: possibility or impossibility
2. There is absolutely no mention of the word "false" in the definition

The first definition (of consistency) makes no mention of the word "false" and so this definition (of inconsistency) makes no mention of it either. We don't care if it is possible for all statements in a set to be false. That's not helpful, that's just a sad state of affairs.

What is helpful is knowing if it is possible that, in some state of affairs, all the statements can be true. If yes, then it is a consistent set. If no, then it is an inconsistent set.

For example, consider this set of statements:

The lights are on.
The door is closed.
Class is in session.

Three statements are in our set. Are they currently all true? Who cares—that's a trivial fact. What matters is if it is possible that all three of these statements can be true. That's what matters.

In this example, the answer is yes. So we know that the set is consistent (not "the same," but consistent). Now consider this set:

George Washington was Russian.
The moon is smaller than the state of Virginia.
Dogs walk on two legs.

We should see that all of these statements are clearly false. So what does that tell us?

A: Nothing.

Knowing that all three statements are indeed false tells us nothing about the set. We want to know if it is possible that all three statements could be true under some state of affairs.

So, might it not be possible that Washington was in fact Russian? Couldn't the moon be super tiny (in moon-terms)? And have you never seen those circus dogs trained to walk on two legs? Could it not be the case that every dog regularly walks as they do? We should be able to see that even though these three statements are false under the current conditions of the world, it is *possible* that they could all be true. Thus, this is a consistent set of statements.

Now consider this set:

The lights are on.

The door is closed.

The lights are not on.

Immediately, we should be able to tell that there is no way, under any conditions at all, for these three statements to be true at the same time. At least one will always end up false.

This example should be easy, but in many cases we may find it difficult to determine if the set of statements is or is not consistent. However, as with the other logical properties, later we will develop tools that help us determine this with considerable ease and accuracy.

Logical Properties of Statements and Sets of Statements

Logical Truth / Logical Tautology: a statement is logically true if and only if it is not possible for it to be false under any conditions

Logical Falsehood / Logical Self-contradiction: a statement is logically false if and only if it is not possible for it to be true under any conditions

Logical Indeterminacy / Logical Contingency: a statement is logically contingent if and only if it is possible for it to be true under some conditions and false under other conditions

Logical Equivalence: a pair of statements are logically equivalent if and only if it is not possible for one to have a different truth value than the other under the same conditions

Logical Consistency: a set of statements is logically consistent if and only if it is possible for all members of the set to be true under the same conditions

Logical Inconsistency: a set of statements is logically inconsistent if and only if it is not possible for all members of the set to be true under the same conditions

Practice Questions

Indicate whether the following statements are true or false.

Justify your answer (in 1 or 2 sentences).

1. Some arguments are false.
2. An argument is any set of statements.
3. All arguments have more than one premise.
4. Some true statements are valid.
5. All valid arguments have true conclusions.
6. Some sound arguments have false premises.
7. Any argument with a true conclusion is sound.
8. All weak arguments have false conclusions.
9. No cogent arguments have false premises.
10. No cogent arguments have false conclusions.
11. No cogent argument is valid.
12. Some sound arguments are invalid.
13. Some valid arguments are unsound.
14. All invalid arguments have logically consistent premises.
15. All arguments made up entirely of true statements are valid and sound.
16. Some valid arguments have inconsistent premises.
17. An argument whose conclusion is logically equivalent to one of its premises must be valid.

2.

FALLACIES

Introduction

Ever since Aristotle helped to clearly express the principles of logic, logicians have tried to develop quick and readily available means to identify errors in reasoning. Some errors in reasoning are easy to identify and some are difficult to spot. Often this is due to rushing to conclusions or listening poorly to what is stated in the premises. When these errors are excessive, we spot them easily.

For example, if our friend is worried about the grade they got on a math exam, they may try to tell you:

> I'm totally going to fail this class. I mean, I just got a C on the midterm exam!

You can probably see they are overreacting to one grade. You likely also see that the grade they are reacting to is not even a failing mark itself. Their mind is leaping to a definitive conclusion on the basis of one premise that cannot support the truth of such a claim with *that much* certainty.

Let's try some other examples. Consider the following two statements:

1. If I have twenty dollars, I will get a music album.
2. I got a music album.

And pause.

Stop for a second and consider those two statements. Take a moment to consider what comes next. Where does your mind leap? (Write it down, we'll come back to it later.)

Let's try another example. Consider the following statements:

1. Marx tried to explain what is wrong with capitalism, the kind of social and economic system we have in the United States.
2. Of course, Marx was a tree-hugger communist who hated Americans.

Again, where does your mind leap? (Write that down too; we'll come back to both of these examples.)

Sometimes, even when we are careful and trying to think clearly, figuring out if our minds have done a good job in drawing conclusions is not easy. This is where many of the tools of logic come into play. Over the last two thousand years a *very* wide range of tools have been developed, too many to cover in this class. However, we should be familiar with two general types of error: formal and informal fallacies.

Why We Study Fallacies

A logical fallacy is simply an error in reasoning. These errors are useful to learn, as they help us identify flaws in our own thinking as well as errors in what other people present to us.

While they may represent "bad" reasoning in many senses, we should not belittle the importance of fallacies. A great deal of bad thinking is actually very convincing. That fact should alarm us. So, I will say it again to make it stick:

A lot of bad thinking is very convincing.

Put differently:

We study fallacies because they work.

Though they measure poorly by standards of reasoning, they are effective in getting people to draw conclusions that are either invalidly concluded or weakly inferred. Studying fallacies empowers us to use our own minds more effectively by not repeating these errors as well as safeguard ourselves (and those close to us) from the influence of these errors.

Formal Fallacies

Some fallacies can be identified simply by virtue of the sentence structure found in an argument. We call this sentence structure "syntax" to recognize that we are talking simply about how statements are logically formed.

Formal Fallacy: an error of reasoning found solely in the syntax (form) of an argument's statements resulting in an invalid argument form.

We will explore this in much more detail in later chapters, but for now we can revisit our first example above.

1. If I have twenty dollars, I will get a music album.
2. I got a music album.

Take a look at the first statement. This has a special kind of syntax, a special kind of structure in how the statement is formed: it is a conditional statement. The structure is pretty easy to recognize:

IF (fill in the blank), **THEN** (fill in the blank).

In seeing this structure, we don't really care what is in the blank spots. We just see that all conditional statements have this important structure:

They have an "if-clause"

They have a "then-clause"

We even see this when the word "then" is not present, as we have in our example. Another way to say this is that even when small changes are made, we can still recognize the same *form* of a conditional statement. "Form" is another way to say "syntax" or "structure."

For a conditional statement, the if-clause has a name: we call it the antecedent. The then-clause is called the consequent.

When one of our premises is a conditional statement, we may find that another premise has a very specific relationship to one of the clauses in the conditional. In our example, the second premise is affirming the truth of the "then-clause." This is a pattern. We can spot this pattern in the raw form of how the argument's premises relate to one another. More importantly, ***the pattern that is emerging is often completed in the leap of the mind*** we make. Many folks will conclude:

"I had twenty dollars."

Put differently, many folks will conclude that the conditional's antecedent (the if-clause) is true. The full pattern is now easy to see:

1. **If** (antecedent), **then** (consequent).
2. The (consequent) is true.

Therefore, the (antecedent) is true.

This is the argument's full pattern (i.e., the argument's full form). This particular pattern is so common that logicians saw fit to name it: we call it "Affirming the Consequent." This pattern is also invalid.

To understand why this argument is invalid, we can produce a counterexample: a demonstration of how the two premises can be true, yet the conclusion false. For example, I say:

1. It is true that **if** I had twenty dollars, **then** I would have gotten a music album.
2. It is true that I got a music album—*it was given to me as a gift because I am broke*.

So, it is false that I had twenty dollars.

The counterexample shows that from these two premises we cannot confidently conclude that I had twenty dollars. This would be a sketchy leap of the mind.

Note that even if we take "I would get a music album" to mean "I bought a music album," we still cannot avoid an invalid inference if we try to reason this way. Consider:

1. It is true that **if** I had twenty dollars, **then** I would have *bought* a music album.
2. It is true that I bought a music album.

So, I had twenty dollars.

Our counterexample could look like this:

1. It is true that **if** I had twenty dollars, **then** I would have *bought* a music album.
2. It is true that I bought a music album—*it was on a half-off sale, which is good because all I had was $10*.

So, it is false that I had twenty dollars.

You might think that you are being duped in these examples, as though the counterexamples are tricks. This is not so. The counterexamples are developed just to help us see one very simple thing: the possibility exists that the premises can be true and yet the conclusion false. That is:

> Counterexamples demonstrate an actual possibility that the argument has all true premises and a false conclusion (i.e., it is invalid).

In every such instance, the counterexample helps us see that we are stretching our mind to claims that lie beyond what a conditional statement can support. In later chapters we will look extensively at conditional statements, for they are notoriously difficult to properly understand.

Affirming the Consequent is a common formal fallacy. Indeed, saying it is "common" is an understatement. Generation after generation, people all over the world have made this exact same mistake dozens, hundreds, perhaps thousands of times. We would do well to remember this formal pattern so we can avoid more of the same error in how our own mind leaps.

Let's look at another common formal fallacy. Consider the following:

> If I go to the store, I'll get some milk.

Pause now. Did you notice how this is once again a conditional statement? As it turns out, many people struggle to understand just what a conditional statement is asserting. Because they misunderstand it, they are prone to make errors in thinking through it.

Continued, we might discover that:

1. If I go to the grocery store, I'll get some milk.
2. I did *not* go to the grocery store.

Q: Where does your mind leap?

Many folks will conclude on the basis of these two statements that:

> "I did *not* get some milk."

If that's what you concluded, you are not alone...nor are you correct. We have no firm basis to claim that we did not get the milk. In short, this is an invalid inference. Here's the formal pattern:

1. **If** (antecedent), **then** (consequent).
2. The (antecedent) is false.

Therefore, the (consequent) is false.

This pattern is called Denying the Antecedent, due to how we rely upon in the second premise to arrive at our conclusion. This is a well-known formal fallacy. Our conclusion is on shaky ground, and a counterexample makes it easy to see why:

1. If I go to the grocery store, I'll get some milk.
2. I did not go to the grocery store.

I did get some milk—*the coffee shop had some*.

The savvy student may start to pick up on what is going wrong in the leap our minds make when we commit such formal fallacies. In both Denying the Antecedent and Affirming the Consequent, we are misunderstanding what the conditional statement really means. We jump the gun, and our mind leaps to a claim that is not supported.

> We *think* the conditional statement will allow such a jump, because we *think* the conditional statement asserts a very strong relationship between the if-clause and the then-clause. *It does not*.

An "if-then" statement makes it look like we are safe to make strong connections between both clauses. Sometimes it does, but not always. Later we will study conditional statements at length and learn how to safely make guaranteed claims based on them. For now, at least we know there are these two decidedly unsafe ways to make claims based on them.

Before moving on to the informal fallacies, we should note that the unsafe inferences we just learned have a special quality to them. They are both based on a misunderstanding of a connection between two statements (i.e., the two clauses). These connections are very important to understand. For example, the "if-then" connection is notoriously tricky—until we understand it, after which it is really quite straightforward. These "connections" between the component statements are what create the "form" of the overall statement. Successfully asserting the truth (or painfully insisting on falsehood) often rests entirely on the form of our overall statements. The content of our claims matters little if we are lost in the connections made between them.

Informal Fallacies

Many errors of reasoning are found in poor use of syntax. In later chapters we will focus our study on syntax to avoid these errors. However, these are not the only kinds of errors regularly found in arguments. There are *other patterns*, patterns found not in the form of the statements, but patterns in *how those statements are used* in an argument. These *non*-formal patterns are called *in*formal fallacies.

Informal fallacies are often described as errors that occur in the content of an argument. So if "form" is like shape, "content" is what fills that shape. We do not need to know the content of the "if-clause" in a conditional statement to see errors like Denying the Antecedent, because it is a formal fallacy. Yet in an important sense, we often do need to know the content of the antecedent in order to see an informal fallacy at play.

I should stress that seeing content alone is not enough. Knowing the meaning of a statement alone does not reveal the pattern of an informal fallacy. For our purposes, we will do better to think more in terms of what one *does* with language. For the *informal fallacies are more akin to what one does with their claims* than *what their actual claims may mean.*

Informal Fallacy (a working definition): an error in reasoning found in the patterns of what is done in an argument, specifically in how the statements are used to create influential strategies with little logical merit.

Why We Study the Informal Fallacies

Informal fallacies are a misuse of influential strategies. They are often committed because, like all fallacies, they WORK. We should not be surprised to see bookshelves filled with authors proposing to teach people how to win friends and influence people. People have complex psychological and emotional lives. We can use words to tap into these driving forces to influence people, to get them to accept conclusions even when we have not given proper support for the truth of those conclusions.

In short, "how to win friends and influence people" is nothing short of manipulation.

While there are additional serious concerns over the ethical quality of these strategies, above all, these are not logically respectable methods. These are the informal fallacies.

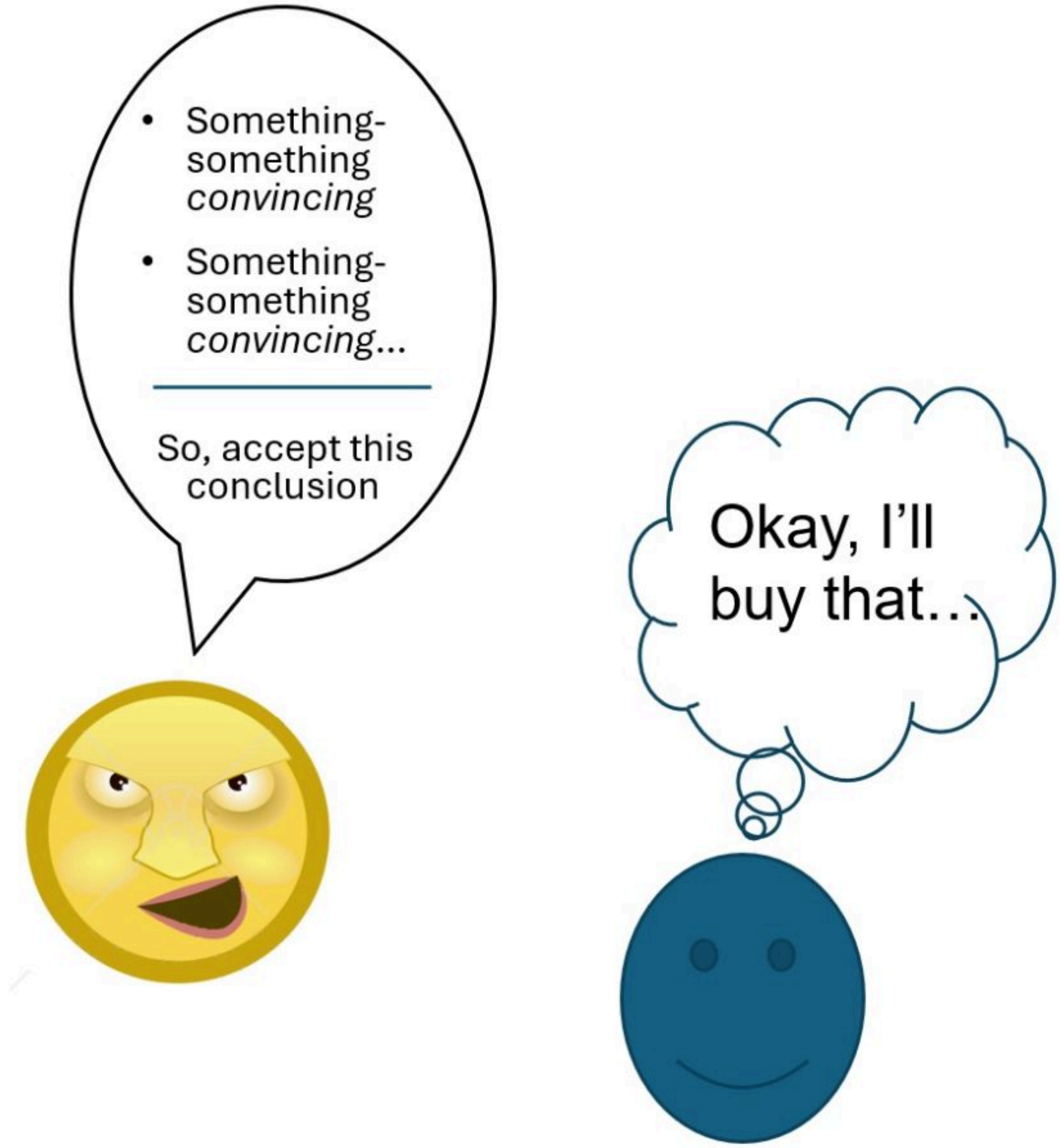
Our approach to the informal fallacies will be to think of them as shady strategies used to convince us of a conclusion. Strictly speaking, these patterns of argument (i.e., "strategies") are frequently committed innocently. Many folks are not aware their arguments follow these patterns, or they are not aware there is something wrong with these patterns. Still, framing the informal fallacies as shady strategies is useful in understanding them and the context in which they are often used.

There are two main characters in this approach:

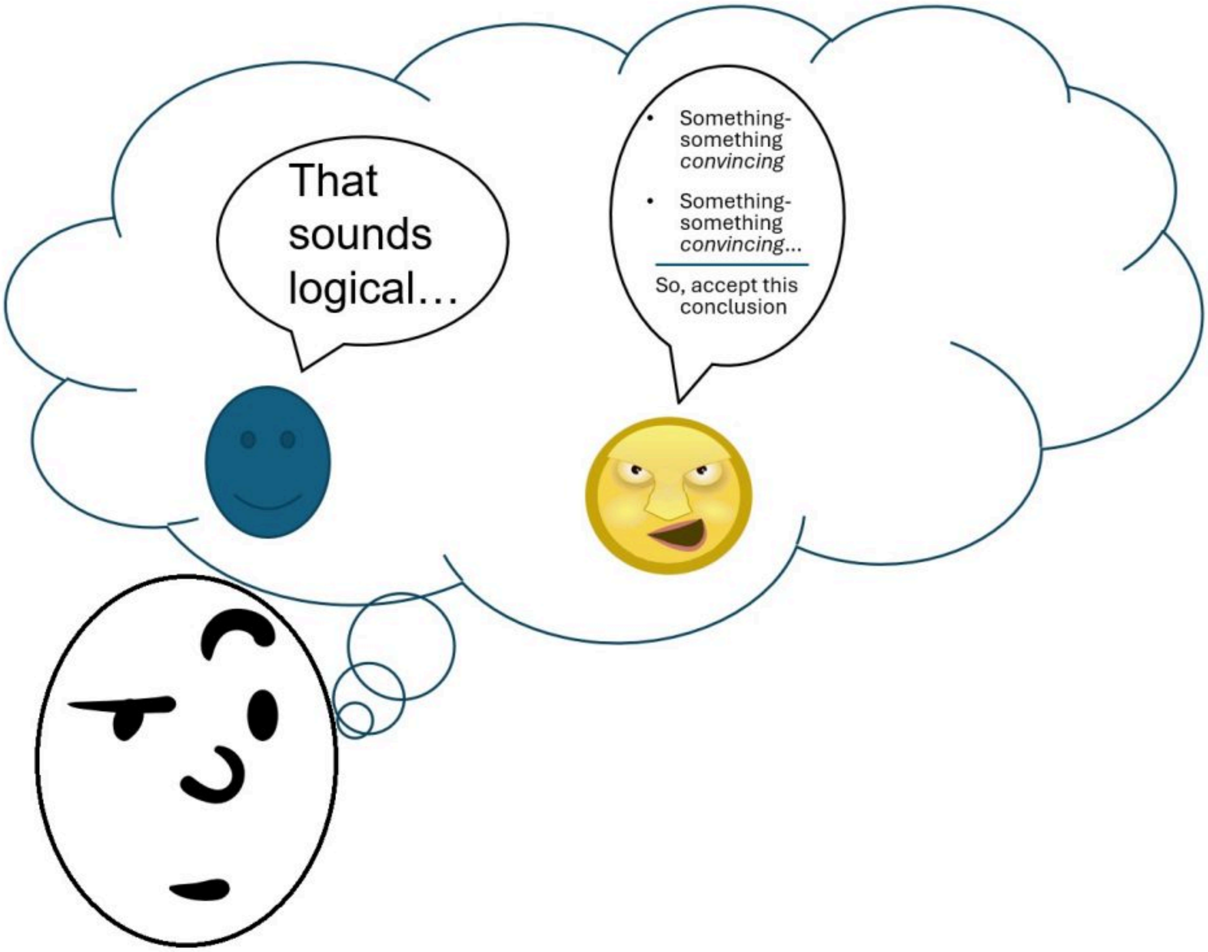
The Author: the person who is giving the argument

The Audience: the person(s) who receives the argument

The author of a fallacy is using some kind of nefarious tactic to land his/her conclusion. They often know that providing a logically compelling argument for their conclusion is difficult. Winning over their audience with the best *respectable* argument *that they can muster* will be too hard for them. So, they take the easy way out. They give an argument which has little respectability but high "convince-ability." If we are not careful, we will fall for their game. They will dupe us like a con artist who plays us for fools (and in truth we would be foolish to accept their arguments).



What makes all of this even more challenging is that **"the author" and "the audience" are sometimes the same person**. As odd as that sounds, if we are prone to using these strategies, we often use them on ourselves. We *can convince ourselves* of things using the same con-man strategies that others would use when they are manipulating us. The informal fallacies can be so powerful that we do not see that we can play ourselves.



Organizing Informal Fallacies

Logicians have identified many informal fallacies, far too many for us to cover at the moment. So, we will look at a few of the more common and troubling ones. This is by no means a complete list, and students should be aware that any list of fallacies reflects a given logician's sense of the errors of reasoning that are most important to cover.

We will group the fallacies we cover into three sets with common qualities. This should help us see the similarities among the fallacies.

- Set I: Errors of Relevance
- Set II: Appeals to Unwarranted Assumption
- Set III: Weak Induction

Keep in mind that all the fallacies in each set share a common quality. For example, all the fallacies in the Relevance set will be variations on a common theme: presenting irrelevant premises. As it turns out, there are many ways in which an author can present irrelevant premises. These ways are strategically different, each using a spin on the general tactic of using irrelevant claims to convince their audience to accept their conclusion. The same is true of the other sets—within the sets, the fallacies are all variations on a common theme.

Additionally, we will often identify fallacies as a "special type" of a broader fallacy. This is done to help the reader understand the important twists they may see in a given strategy.

Table of Informal Fallacies

SET I: Errors of Relevance

Ad Hominem

- Hypocrisy
- Genetic Fallacy

Straw Man

Appeal to Consequences

- Appeal to Desire
- Bandwagon (Appeal to the People)
- Appeal to Force

Appeal to Pity

Red Herring (Avoiding the Issue / Evasion)

- Smokescreen

SET II: Appeals to Unwarranted Assumption

Begging the Question

Inappropriate Authority

False Dilemma (False Dichotomy)

Loaded Question

Gaslighting

SET III: Weak Induction

Hasty Generalization

SET I

Errors of Relevance

What all fallacies of relevance have in common is that they make an argument or response to an argument that is irrelevant to that argument. These strategies seek to distract the audience from subjects or claims that are relevant to the argument at hand.

Fallacies of relevance can be psychologically compelling, but it is important to distinguish between rhetorical techniques that are psychologically compelling, on the one hand, and rationally compelling arguments, on the other. What makes something a fallacy is that it fails to be rationally compelling once we have carefully considered it. That said, arguments that fail to be rationally compelling *may still be* psychologically or emotionally compelling.

Ad Hominem

"Ad hominem" is a Latin phrase that can be translated into English as the phrase "against the man." In an ad hominem fallacy, instead of responding to (or attacking) the argument a person has made, one attacks the person him or herself. In short, one attacks the person making the argument, rather than the argument itself.

The "ad hom" is one of the most powerful (and thus, troubling) fallacies there is, in large part because *people like to think about people*. One might even say that as social creatures we are deeply hardwired to think about other people. We are not nearly so hardwired to think about abstract ideas or logical relationships. So, our thoughts are *easily led* to what we prefer thinking about—people.

There are a few different subvariants of this strategy. Most classically, the ad hominem fallacy attempts to discredit an argument because the source of that argument (i.e., the author of it) is a bad person. Importantly, the author of this fallacy rarely aims this argument *at that* person. In other words, this is not a strategy typically used in direct debate between two parties. The ad hom's intended audience is usually someone else who is not present at the time. The goal of the strategy is to get *the audience* to reject "the bad person's argument" simply because they are a bad person. For example:

> "Look Mary, I know Mike likes to talk a lot about his faith and why he's so convinced his religious beliefs are true. But really now, we can't take anything he says seriously. He's just another religious nutjob who thinks he's found God."

Notice how the author of this argument uses the main elements of this fallacy:

a. Makes it clear that he is aware that *Mike has made arguments* for his beliefs
b. Is not engaging with those arguments themselves
c. Tries *to convince Mary* to *reject those arguments* based on insults to Mike's intellect and integrity

Ad hominem arguments bank on the audience being unwilling to accept things that "bad people" say. So, the author of an ad hominem will typically highlight some aspect of the person's character they think they can paint in a negative light or that they expect their audience to grasp on their own.

For example, suppose that Hitler had constructed a mathematical proof in his early adulthood (he didn't, but just suppose). A contemporary author of an ad hominem fallacy probably does not need to say anything bad about Hitler. They will rely on their audience to already believe that Hitler = bad. Yet the validity of Hitler's mathematical proof should stand on its own; the fact that Hitler was a horrible person has nothing to do with whether the proof is good. Likewise with any other idea: ideas must be assessed on their own merits, and the origin of an idea is neither a merit nor demerit of the idea.

Of course, the author may find other ways to zero in on a person's negative traits. So, there are variants of this nefarious strategy that use:

> A person's conduct: See the Hypocrisy (Tu quoque) fallacy
> A person's situation: See the Genetic fallacy

We will look at these next, but keep in mind the general structure of all ad hominem strategies:

- Being aware of an argument made by someone else
- Avoiding evaluating the other person's argument
- Calling out something negative about the origin of that argument
- Using those negative things as a basis *to reject* the argument

This last element is important to keep in mind. An ad hominem *strategy* has a purpose—to convince the audience to reject an argument. If an author is simply calling someone names, belittling them, or otherwise being mean-spirited in their reference to them, but they are *not trying to use that* as the support for rejecting their argument, then they are not committing an ad hominem fallacy. Their bad language, crude behavior, or even immorality may just mean they are a mean person. Yet even mean people can muster quality arguments.

Hypocrisy (a.k.a. Tu quoque)

"Tu quoque" is a Latin phrase that can be translated into English as "you too" or "you, also." The basic strategy in this Hypocrisy fallacy is again to ignore the argument given by others in favor of looking at some aspect of the person making that argument. In this case, the author of a hypocrisy fallacy looks at what another person *does* rather than evaluate their argument. Their character is not used against them directly; instead, their conduct is used against them.

The tu quoque fallacy is used in two ways:

> First: as a way of avoiding answering a criticism by bringing up a criticism of your opponent rather than answering the criticism.
>
> Second: as a way of avoiding evaluation of an argument made by others by bringing up their own conduct as evidence that their argument should not be taken seriously.

Let's look at the first way the hypocrisy fallacy is used.

For example, suppose that two political candidates, A and B, are discussing their policies and A brings up a criticism of B's policy. In response, B brings up her own criticism of A's policy rather than respond to A's criticism of her policy. B has committed the tu quoque fallacy. The fallacy is best understood as a way of avoiding having to answer a tough criticism that one may not have a good answer to. This kind of thing happens all the time in political discourse. This is also a specific case in which the author's intended audience (of the basic ad hominem strategy) is the very same person whose argument is rejected.

Tu quoque, as I have presented it, is fallacious when one raises a criticism simply to avoid having to answer a difficult objection to one's argument or view. However, there are circumstances in which a tu quoque response is not fallacious. So, be careful in judging someone this way. Let's look at a situation in which it might look like a nefarious tactic is being used, but in reality is not.

If the criticism that A brings toward B is a criticism that equally applies not only to A's position but *to any position*, then B is right to point this fact out. For example:

> Suppose that Allen criticizes Billy for taking money from special interest groups. In this case, Billy would be totally right to respond that not only does Allen take money from special interest groups, but *every political candidate* running for office does. Billy is not accusing Allen of hypocrisy. He is pointing out that his own conduct is not unique (it's the same as Allen and everyone else's conduct) and thus not a basis for critique. That is just a fact of life in American politics today.

So *Allen really has no criticism at all of Billy* since everyone does what Billy is doing and it is in many ways unavoidable. Thus, Billy could (and should) respond with a "you too" rebuttal, and in this case, that rebuttal is not a tu quoque fallacy.

Now let us look at the second use of the Hypocrisy fallacy.

Here is an anecdote that reveals an ad hominem fallacy.

> In an ethics class, students read the work of Peter Singer. Singer had made an argument that it is morally wrong to spend money on luxuries for oneself rather than give all of your money that you don't strictly need away to charity. The essence of the argument is that there are every day in this world

children who die preventable deaths, and there are charities who could save the lives of these children if they were funded by individuals from wealthy countries like our own.

In response to Singer's argument, one student in the class asked:

"Does Peter Singer give his money to charity? Does he do what he says we are all morally required to do?"

The implication of this student's question was that if Peter Singer himself doesn't donate all his extra money to charities, then his argument isn't any good and can be dismissed. But that would be to commit the ad hominem fallacy of Hypocrisy. Instead of responding to the argument that Singer made, this student attacked Singer himself. That is, they wanted to know how Singer lived and whether he was a hypocrite or not. Was he the kind of person who would tell us all that we had to live a certain way but fail to live that way himself? But all of this is irrelevant to assessing Singer's argument.

To see this, we only need to rationally evaluate the student's proposal. Suppose that Singer didn't donate his excess money to charity and instead spent it on luxurious things for himself. Even if it were true that Peter Singer was a total hypocrite, his argument may nevertheless be rationally compelling. And it is the quality of the argument that we are interested in, not Peter Singer's personal life and whether or not he is hypocritical.

Whether Singer is or isn't a hypocrite is irrelevant to whether the argument he has put forward is strong or weak, valid or invalid. As a good logician, the student should be focused on these qualities of the argument. The argument stands on its own and it is that argument rather than Peter Singer himself that we need to assess.

Nonetheless, many of us still find something psychologically compelling about the question: Does Peter Singer practice what he preaches? I think what makes this question seem compelling is that humans are very interested in finding "cheaters" or hypocrites—those who say one thing and then do another. We *want* to think about people. That said, whether or not the person giving an argument is a hypocrite is irrelevant to whether that person's argument is good or bad.

As before, we need to be careful in judging someone's argument as an example of this fallacy. Not every instance in which someone attacks a person's character is an ad hominem fallacy. Suppose a witness is on the stand testifying against a defendant in a court of law. When the witness is cross-examined by the defense lawyer, the defense lawyer tries to go for the witness's credibility, perhaps by digging up things about the witness's past.

For example, the defense lawyer may find out that the witness cheated on her taxes five years ago or that the witness failed to pay her parking tickets. This isn't an ad hominem fallacy because in this case, the lawyer is trying to establish whether what the witness is saying is true or false. In order to determine that, we have to know whether the witness is trustworthy. These facts about the witness's past may be relevant to determining whether we can trust the witness's word. In this case, the witness is making claims that are either true or false rather than giving an argument.

In contrast, when we are assessing someone's argument, the argument stands on its own in a way the witness's testimony doesn't. In assessing an argument, we want to know whether the argument is strong or weak and we can evaluate the argument using the logical techniques surveyed in this text.

In contrast, when a witness is giving testimony, they aren't trying to argue anything. Rather, they are simply making a claim about what did or didn't happen. So although it may seem that a lawyer is committing an ad hominem fallacy in bringing up things about the witness's past, these things are actually relevant to establishing the witness's credibility.

Genetic Fallacy

The genetic fallacy is a variant of the basic ad hominem fallacy. This fallacy occurs when one argues (or, more commonly, implies) that the origin of something (e.g., a theory, idea, policy, etc.) is a reason for rejecting (or accepting) it.

For example, suppose that Jack is arguing that we should allow physician-assisted suicide and Jill responds that that idea first was used in Nazi Germany. Jill has just committed a genetic fallacy because she is implying that because the idea is associated with Nazi Germany, there must be something wrong with the idea itself. What she should have done instead is explain what, exactly, is wrong with the idea, rather than simply assuming that there must be something wrong with it since it has a negative origin. The origin of an idea has nothing inherently to do with its truth or plausibility.

The basic strategy of this variant is to hide a straight-up attack on a specific individual who came up with an idea in favor of a generic reference to the group or the time period in which an idea developed. Of course, we know that ideas do not grow on trees. So, this strategy always implicitly relies on calling out some person(s) who developed the idea—and they are bad people, so this is a bad idea.

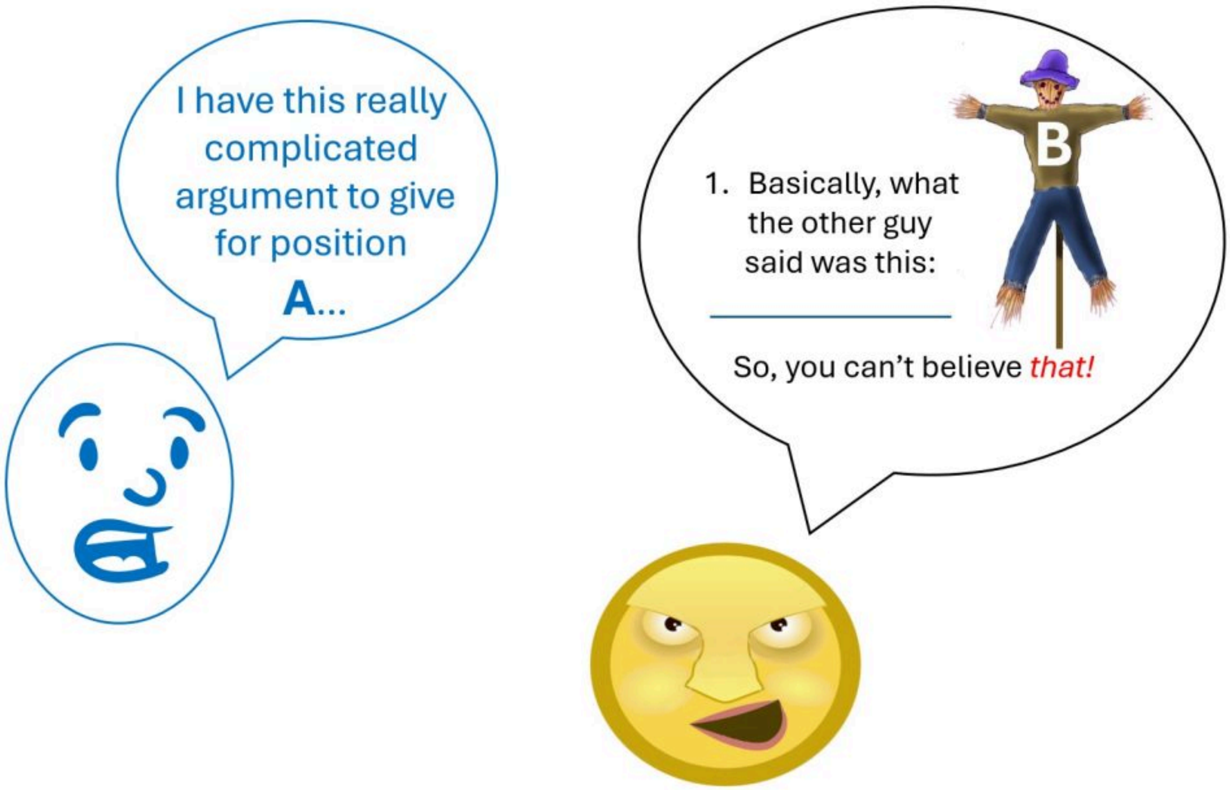
Although genetic fallacies are most often committed when one associates an idea with a negative origin, it can also go the other way: one can imply that because the idea has a positive origin, the idea must be true or more plausible.

For example, suppose that Jill argues that the Golden Rule is a good way to live one's life because the Golden Rule originated with Jesus in the Sermon on the Mount (it didn't, actually, though Jesus does state a version of the Golden Rule). Jill has committed the genetic fallacy in assuming that the (presumed) fact that Jesus is the origin of the Golden Rule has anything to do with whether the Golden Rule is a good idea.

Straw Man

Suppose that my opponent has argued for a position, call it position A, and in response to his argument, I give a rationally compelling argument against position B. That sounds like a foolish thing to do! After all, my opponent is arguing for position A. So why did I argue against position B?

Answer: Because position B is *related to* position A, but is much less plausible (and thus much easier to refute). What I have just done is attack a straw man—a position that "looks like" the target position, but is actually not that position.



Straw men are easy to knock down. When one attacks a straw man, one commits the straw man fallacy. The straw man fallacy misrepresents one's opponent's argument and is thus a kind of irrelevance.

Here is an example:

Two candidates for political office in Colorado, Tom and Fred, are having an exchange in a debate. Tom has laid out his plan for putting more money into health care and education. Fred has laid out his plan, which includes earmarking more state money for building more prisons, which will create more jobs and thus strengthen Colorado's economy. Fred responds to Tom's argument that we need to increase funding to health care and education as follows:

> "I am surprised, Tom, that you are willing to put our state's economic future at risk by sinking money into these programs that do not help to create jobs. You see, folks, Tom's plan will risk sending our economy into a tailspin, risking harm to thousands of Coloradans. On the other hand, my plan supports a healthy and strong Colorado and would never bet our state's economic security on idealistic notions that simply don't work when the rubber meets the road."

Fred has committed the straw man fallacy. Just because Tom wants to increase funding to health care and education does not mean he does not want to help the economy. Furthermore, increasing funding to health care and education does not entail that fewer jobs will be created. Fred has attacked a position that is not the position that Tom holds, but is in fact a much less plausible, easier to refute position. However, it would be silly for any political candidate to run on a platform that included "harming the economy." Presumably no political candidate would run on such a platform. Nonetheless, this exact kind of straw man is ubiquitous in political discourse in our country.

Here is another example.

Nancy has just argued that we should provide middle schoolers with sex education classes, including how to use contraceptives, so that they can practice safe sex should they end up having sex. Fran responds:

> "Proponents of sex education try to encourage our children to adopt a sex-with-no-strings-attached mentality, which is harmful to our children and to our society."

Fran has committed the straw man (or straw woman) fallacy by misrepresenting Nancy's position. Nancy's position is not that we should encourage children to have sex, but that we should make sure that they are fully informed about sex so that if they do have sex, they go into it at least a little less blindly and are able to make better decisions.

As with other fallacies of relevance, straw man fallacies can be compelling on some level, even though they are irrelevant. It may be that part of the reason we are taken in by straw man fallacies is that humans are prone to "demonize" the "other"—including those who hold a moral or political position different from our own.

This tendency to "demonize the other" dovetails well with other errors of reasoning, such as the whole range of ad hominem strategies. If we are prone to demonize people who are different from us, we are ready and eager to believe that *they must hold extreme views* or morally outrageous beliefs. Poor reasoning breeds ever more poor reasoning. It becomes easy to think bad things about those with whom we do not regularly interact. And it is easy to forget that people who are different than us are still people just like us in all the important respects—especially in their ability to offer quality arguments.

Appeal to Consequences

Logic encourages us to maintain some overarching standards. For example, our conclusions should rest on relevant beliefs that do a good job of supporting the truth of those conclusions. We should strive to meet that standard.

However, all too often we muddy this standard by forgetting that our support should be relevant *to the truth* of our conclusions, not simply relevant to something *important to us*. Many things are *of consequence to us* in some general way. Sometimes it is hard to distinguish what is important to us from what is relevant to the truth of a claim.

We often have things that are of personal interest to us, and because they are valued *that way* we tend to think they are "relevant" to all sorts of things. This slip of the mind is a failing we should avoid. The fact that something may have a beneficial consequence to us does not automatically mean it is (a) true or (b) relevant to the truth of another claim.

There are two broad ways the consequences of a belief may interest us personally:

- Positive ways: because it benefits us
- Negative ways: because we would rather avoid it

Simply put, *we are attracted to desirable things, and we are averse to painful or frightening things*. This is the human condition, but it often has little to do with the truth of specific claims.

We'll first look at the broadest form of Appeal to Consequences arguments. Then we will look at three specific ways this strategic error plays out in its positive and negative manners.

The Appeal to Consequences fallacy consists of the mistake of trying to assess the truth or reasonableness of an idea based on the (typically negative) consequences of accepting that idea.

For example, suppose that the results of a study revealed that there are IQ differences between different races (this is a fictitious example; there is no such study that I know of). In debating the results of this study, one researcher claims:

> If we were to accept these results, it would lead to increased racism in our society, which is not tolerable. Therefore, these results must not be right, since if they were accepted, it would lead to increased racism.

The researcher who responds in this way has committed the Appeal to Consequences fallacy. Again, we must assess the study on its own merits. If there is something wrong with the study, some flaw in its design, for example, then that would be a relevant criticism of the study. However, the fact that the results of the study, if widely circulated, would have a negative effect on society is not a reason for rejecting these results as false. The consequences of some idea (good or bad) are irrelevant to the truth or reasonableness of that idea.

Note that in our example the researchers, being convinced of the negative consequences of the study on society, might rationally choose not to publish the study (for fear of the negative consequences). This is totally fine and is not a fallacy (we often see this kind of argument made in public health decisions, where the impact on public behavior is central to their conclusions). The fallacy consists *not in choosing not to publish* something that could have adverse consequences, but *in claiming that the results themselves* are undermined by the negative consequences they could have.

This just goes to show that the consequences of an idea are irrelevant to the truth or reasonableness of an idea. Now we will look at three unique spins on this general error, starting with two variants that stress positive consequences.

Appeal to Desire

When we have a positive view of something, we tend to want whatever is around it too. This is because we associate those things with the main thing we think is positive in the first place. Think of all the ads with sexy models shilling for cars or beers or whatever. What does sexiness have to do with how good a beer tastes? Nothing. The ads are trying to engage your emotions to get you thinking positively about their product. The strategy here is for the author of the argument to link their conclusion (e.g., "You should buy this product.") to something beneficial for you ("You will be seen as sexy or desirable.").

You might think that this is a pretty flat-footed approach. After all, in a calm moment of clear reflection, we can easily see that being seen as sexy has no relevance to the truth of this car's quality, or that beer's taste, or this vacation spot's fun opportunities, and so forth. Yet think for a moment how often you see this done. Advertisers will spend millions on such campaigns because, as we have seen, *they work*.

Flat-footed or not, we are apt to be duped by these strategies when they are used without being in our faces about it. We see the ads make explicit claims that do not reference sexiness, or wealth, or whatever else they are displaying in the background. The Appeal to Desire strategy is most effective on us when we merely sniff it in the atmosphere of an argument. This makes thinking explicitly about it difficult, because it is not at the front of our consciousness.

Another way to pull this off is to link the conclusion to things that comfort us. The nefarious author is apt to use things they know we identify with (or things that are familiar to us) to make the conclusion more appealing. I might have fond memories of *momma's home cooking*, but it is unlikely that this box of powdered mashed potatoes really does taste like that. I might be nostalgic about a fictional time in U.S. history when "*Americans settled the Wild West*," but it is unlikely that buying this work shirt will bring me closer to becoming the rugged individual those stories promoted.

Bandwagon (a.k.a. Appeal to the People)

Another positive variant of the Appeal to Consequences fallacy uses base desires that are deeply rooted in our social nature. An extremely common technique, especially for advertisers, is to appeal to people's underlying desire to fit in, to be hip to what everybody else is doing, to not miss out. This is the bandwagon appeal.

Think of how many times you have seen an advertisement that assures us that a certain television show is #1 in the ratings—with the tacit conclusion being that we should be watching, too. But this is a fallacy. We've all known it's a fallacy since we were little kids: the first time we did something wrong because all of our friends were doing it, too, and our moms asked us, "If all of your friends jumped off a bridge, would you do that too?" Still, now we want to watch that show...the strategy still works on us.

Note that the nefarious author who uses a bandwagon strategy is relying on a powerful psychological need they know their audience harbors. Humans are social animals. We want to feel a sense of belonging to some group. We desire to fit in and are averse to feeling like we don't belong.

We all likely remember our early teen years when it was extremely important to us that we belonged to some group in school. We did not all want to be one of the popular kids, but we all wanted to have some sense of shared identity. Perhaps we found it in a small group of "outsiders." We may have been "misfits" to the broader community, but within our group it was important that we fit in.

So, the group doesn't always have to be a large group—it could be an exclusive group, just so long as it is a group we find value in. Expressing positive value in group membership can be subtle. The nefarious author doesn't have to come out and say, "*this group is great!*" They just have to trigger the audience to keep in mind that they are (or want to be) part of a group. Make no mistake, a simple reminder that "everyone is watching this" is enough to get the job done. After all, "everyone" is a good group—it's big, it's safe, and it has the appeal of common sense. The "must-see show of the season" triggers the sense that this is what everyone is watching, and that's exactly the message the nefarious author needs us to hear. We will conclude that we too should be watching this show, because somewhere in the back of our head it makes sense to fit in. The "must-see" phrase helps us understand that the broader social group expects this of us.

Worth noting is an advertiser's favorite tactic of using *small* elite groups to hock their wares. We may not want to be like everyone else. Many of us favor the idea that we do not fit in with the crowd—because we are special and unique and authentic people...so we'll buy the leather jacket only *the fashionistas* appreciate (never mind that it is overpriced and was made with exploited labor). We will also show our belonging with elite groups by buying from those who have cultivated their brand as elite. The luxury car we drive is "good" because wealthy people drive it too (never mind the poor reliability scores or low customer satisfaction ratings). We want to be seen *as one of them*. We are still jumping on a bandwagon; it just happens to be a small one.

Appeals to hypermasculinity or toughness offer the same opportunity to identify as a rare breed. "*It's a Jeep thing—if you have to ask, you wouldn't understand*." After all, only the rough and rugged understand the true value of how the adventurous live. For the nefarious author this has the double benefit of shielding them from critique. My conclusions are legitimate because they are the conclusions of this group, AND anyone who offers counterclaims cannot do so sensibly without betraying their non-membership with the group.

Of course, let's not forget that we all have strong interests in belonging to multiple groups. So nefarious authors (in this case, advertisers) often use overlapping groups to promote the tacit conclusions of their ads (buy this product). When a celebrity like John Travolta wears a watch as he walks towards his jet aircraft, we want that watch! He's sexy, he's rich, he's adventurous, and he's famous. So if you can break into *that crowd*, then you know you've made it too. The watch is way too big for your wrist, but no matter. You still want it.

Appeal to Force

A negative variant of the Appeal to Consequences uses our fear of bad things to get us to accept the truth of some claims. In this fallacy the nefarious author attempts to convince their audience to believe something by threatening them. Threats pretty clearly distract one from the business of dispassionately appraising premises' support for conclusions, so it's natural to classify this technique broadly as an Error of Relevance.

There are many examples of this technique throughout history. In totalitarian regimes, there are often severe consequences for those who don't toe the party line (see George Orwell's *1984* for a vivid, though fictional, depiction of the phenomenon). The Catholic Church used this technique during the infamous Spanish Inquisition: the goal was to get non-believers to accept Christianity; the method was to torture them until they did.

While these examples suggest that the Appeal to Force strategy is always heavy-handed, most instances are actually quite subtle. The trouble is that people generally don't like being threatened. So, when they are directly threatened, unless that threat is both believable and extreme, they are likely to reject the strategy used on them. Simply put, we resist. For the nefarious author, this is not a good tactic. Better for them to use a *subtle* tactic to pull off this strategy.

For example, your boss seems concerned about you. I mean, after all, you're up for review soon and maybe your promotion or even position in the job is on the line. "I just want to make sure you're going to be in good shape with the company. We'd hate to have to lose you. Let's talk about that triple-overtime I told you about yesterday."

Note that the boss isn't outright threatening you. Indeed, your boss uses passive language. *She* won't fire you—"*we'd have to lose you*," as though it just might be a thing that happens. She is concerned; she is worried for your well-being. That's the official message uttered out loud. But not really. The subtle message is quite the opposite. Real-world appeals to force are often done within contexts that establish the threat.

Appeal to Force strategies are often done with sympathetic words. A nefarious author may appear to want what is good for you. Consider the tone of voice used when people say things like this:

> "I just couldn't bear the thought of what would happen to you if you left me. *I love you more than anything else in the world*."
>
> "Accepting this religious claim is your only salvation, your only hope *to truly find the peace you are looking for*. Without it your eternal soul will be lost."
>
> "I need to know you'll be safe and able to stand on your own two feet. You have to major in finance or pharmacy. We can't keep paying for you to throw your life away on this poetry thing. *Your future is all we care about*."

Most folks who offer arguments like this sound nice. At least, they seem deeply concerned with our well-being. However, the quality of this relationship and my decision to stay in it should be determined by how I might see the relationship, not the consequences of what my partner will do (to me) if I should leave. Likewise, our well-being doesn't *make* those religious claims true—what might happen to me if I don't accept those claims offers no support for the truth of those claims. Last, as caring as my parents might come across, their threat to withhold financial support for my schooling is not relevant to whether it is true that I should select this or that major. These are all just sugar-coated threats.

Last, when we look at the conclusions of these arguments, we can see, at least in principle, that there are relevant reasons to accept them—an author could have spoken directly to the truth of those claims and given us direct reasons to accept them. However, they opted instead not to address them. They could have spoken to why this relationship is worth salvaging, or why the beliefs of this religious tradition are true, or why these majors make sense for us to pursue. None of these were mentioned. Instead they made assorted threats about the consequences we would face if we didn't accept their conclusions.

Appeal to Pity

Another emotional response we are likely to have is pity for others. This is generally a good thing, but as we have seen, a nefarious author can take advantage of their audience by using this to convince them to accept a conclusion without proper support. Consider this example:

> *But Professor, I tried really hard on the exam. I need this class to graduate. If I don't pass, I'm going to be held back. I worked my fingers to the bone! You just have to pass me.*

Note how the error of reasoning here relies heavily on the nature of the appeal. This strategy involves appealing to the pity itself as the support for the truth of a conclusion. This is irrelevant.

Of course, there are *relevant reasons* a professor should pass a student—such as strong performance on the exams, completion of assignments to a satisfactory level, and attendance and participation in class (if this is part of the course's grading structure). However, this author avoids bringing these up. Instead, they invoke their effort, their graduation requirements, and their academic progress. All of these are irrelevant to the truth of whether or not they earned a passing grade. So why bring it up? Answer: to make the professor feel pity for the author. This nefarious strategy is simple: You should accept my conclusion because you feel bad for me.

The recursive nature of the appeal is important. If an argument is made and the claims within it *happen to make* the audience feel bad for some third party, this is **not** an appeal to pity strategy. We only use the audience's pity-response to identify a nefarious strategy when:

a. That response was the whole point of raising these claims (i.e., it was the author's intention to invoke pity)
b. The substance of what invokes pity is irrelevant to the truth of the conclusion
c. The pity is intended for the author of the argument (i.e., it is recursive)

So consider the following commercial:

> For some dogs, Christmas is a terrible time. Left out in the cold, these poor souls have no hope. They suffer in the cold, underfed and unloved. For just $9.99 you can help those who cannot help themselves. Won't you consider supporting our work to rescue these animals?

These sorts of commercials almost always show miserable dogs kept in the cruelest conditions. What sort of person would see this and not feel some pity for those animals? We can expect that most audience members would feel bad for the dogs. Yet we should not jump the gun and claim we are being manipulated by a nefarious author who is trying to dupe us. After all, the conclusion *is* that we should help the dogs (or help the organization whose work is to help the dogs). If the dogs were supremely trained actors and they were not really in dire straits, then we would feel duped. We would feel this way precisely because their miserable condition is actually relevant to the conclusion that we should help alleviate that condition. We're being presented with the images/claims of their suffering because the substance of their suffering is relevant to the conclusion. This removes condition (b) above.

Now consider condition (c) above: is the intent of the commercial to make you feel bad for the organization asking for your help? No. The organization (as the author of the argument) does not want you to feel pity *for them*. There's a third party in this argument (the dogs), and the author's premises suggest cause for feeling bad about *those others* who are in a bad situation. So condition (c) is not met.

We might be concerned about the commercial's intention; that is, condition (a) above may still apply. So ask yourself, is the whole point of the images/claims *just* to make us feel pity? This is a bit tricky to evaluate. When you hear the music played, the tone of the announcer, the terrible images, you certainly think: if nothing else, they are trying pretty hard to make you to feel pity for the dogs. So there is something there. However, without condition (b) making the substance of the dogs' condition irrelevant, we might charitably say the author has little chance of avoiding invoking pity. I suppose they could not tell you about the dogs, but that wouldn't help establish any conclusion to help the dogs. So that seems unreasonable. What's not unreasonable is that they could have left out the sad, desperate music, and they could have conveyed their message without the anguished tone of voice. At the very least, we can temper our concerns by noting that condition (c) is also not met. The pity is almost inevitable (thus, not directly appealed to) and there is no effort to appeal to the pity on behalf of the author. Put differently, some commercials are troubling (from a logical point of view), but they are not outright nefarious.

Red Herring / Avoiding the Issue / Evasion

This fallacy initially gets its name from the actual fish. When herring are smoked, they turn red and are quite pungent. Stinky things can be used to distract hunting dogs, who of course follow the trail of their quarry by scent; if you pass over that trail with a stinky fish and run off in a different direction, the hound may be distracted and follow the wrong trail.

Whether or not this practice was ever used to train hunting dogs, as some suppose, the connection to logic and argumentation is clear. One commits the red herring fallacy when one attempts to distract one's audience from the main thread of an argument, taking things off in a different direction. The diversion is often subtle, with the detour starting on a topic closely related to the original—but gradually wandering off into unrelated territory.

The tactic is also often (but not always) intentional: one commits the red herring fallacy *because* one is not comfortable arguing about a particular topic on its merits, often *because* the author knows their own case is weak; so instead, the arguer changes the subject to an issue about which he feels more confident, makes strong points on the new topic, and pretends to have won the original argument. They have effectively avoided the original issue or evaded the criticisms of their stance.

Note that people often offer red herring arguments unintentionally, without the deceptive motivation to change the subject. Sometimes a person's response will be off topic, apparently because they were confused for some reason. Such cases are usually easily resolved when the audience points this out. When unintentional, the strategy of a red herring is abandoned quickly because the non-nefarious author does not really want to use such a poor strategy, but realizes that this is what they are doing. In the intentional cases, this is not so. The author is much more likely to resist the claim that they are off topic because they know they are *relying on* that strategy to make their case.

People may also be unaware that they are using this strategy even when they are intentionally using tactics that produce it. That is, many people believe they should bring up certain points (and so, they do it on purpose) simply because a certain topic has been raised. Two points are important here:

- They won't admit that they are *trying to evade* the specific subject of an argument (because they are not aware that they have missed the point[1]).
- However, they are perfectly aware that they are *bringing in a new topic* to the discussion or debate (they

simply feel it is appropriate to do so).

These "*Well what about…*" insertions are common when people feel their own beliefs are challenged (especially around political affiliation and moral accusations). For example:

> Caroline may be upset at her husband's wandering eye lately and calls him out on it. Manny instinctively (and thus, unintentionally) responds, "*Well, what about you? You dated plenty of guys in your time before we got married!*"

Manny may not realize he is getting off track (his own behavior is at stake, and it is this current conduct that is central to the accusation). Yet he is very aware that he is letting his wife know he expects this topic to be brought into the conversation. The tactic is intentional; the broader strategy is not.

What-about-isms are convenient ways to hide the shortcomings of one's own position as well as remain oblivious to the irrelevance of what one wants to introduce into the argument. This is often accomplished in the name of an important principle: **don't be a hypocrite**. In this example, Manny expects his wife to follow this principle. So, he feels entitled to introduce her own behavior—as that appears to him to be relevant to the accusation she levied against him. While it is generally a good principle to follow, it belongs in a different argument. If there is something wrong with Caroline's current or past behavior, this judgment stands on its own and has no bearing on whether or not Manny's behavior today is morally culpable. These would be best addressed in two different arguments (as would the *third* argument regarding the accusation that Caroline is hypocritical in her expectations of her husband).

What-about-isms are powerfully effective because they directly challenge the audience to consider a new claim. The audience needs to be clear-minded to identify the irrelevancy and steadfast to remain on topic. This is difficult to do in heated debate, such as we find in political discussions. For example:

> At the neighborhood Halloween party, Rob's liberal-leaning neighbor, Linda, may show dismay at his support of Donald Trump. Linda might ask (and might even ask with genuine curiosity or wonder), "*How could you support a guy who treats women the way he does?*" To which Rob remarks, "*Well what about Clinton? Ol' Bill was having sex in the Oval Office!*"

This might seem like a fair application of the hypocrite principle, except that Bill Clinton's behavior is not the subject of the inquiry. Clinton isn't running against Trump—thus, his conduct is irrelevant. Additionally, if pressed, Linda may express her disgust at Clinton's behavior as well—thus defusing the alleged hypocrisy. However, what-about-isms help a nefarious author feel better about their insertion, as though they have taken the moral high ground (or at least, they have not ended up in a moral underground).

---

1. Many may regard this as similar to an error called "Missing the Point" (Ignoratio Elenchi), which reflects a widely recognized general error of confusion or lack of attention on the part of the author. While this is a malady of good reasoning, I prefer not to treat it as a proper informal fallacy at all (contrary to long-standing tradition). My reason is that "Sorry, I just wasn't listening" involves no real strategic nor tactical error. People make mistakes. We should properly call that "Being Human." If this distinction is not satisfactory to my reader, I would advise classifying Ignoratio Elenchi as an innocent form of Red Herring.

Politicians use the general version of the red herring fallacy all the time. Consider a debate about Social Security—a retirement stipend paid to all workers by the federal government. Suppose a politician makes the following argument:

> We need to cut Social Security benefits, raise the retirement age, or both. As the baby boom generation reaches retirement age, the amount of money set aside for their benefits will not be enough cover them while ensuring the same standard of living for future generations when they retire. The status quo will put enormous strains on the federal budget going forward, and we are already dealing with large, economically dangerous budget deficits now. We must reform Social Security.

Now imagine an opponent of the proposed reforms offering the following reply:

> Social Security is a sacred trust, instituted during the Great Depression by FDR to ensure that no hard-working American would have to spend their retirement years in poverty. I stand by that principle. Every citizen deserves a dignified retirement. Social Security is a more important part of that than ever these days, since the downturn in the stock market has left many retirees with very little investment income to supplement government support.

The second speaker makes some good points, but notice that they do not speak to the assertion made by the first: Social Security is economically unsustainable in its current form. It's possible to address that point head on, either by making the case that in fact the economic problems are exaggerated or nonexistent, or by making the case that a tax increase could fix the problems. The respondent does neither of those things, though; he changes the subject, and talks about the importance of dignity in retirement. He's likely more comfortable talking about that subject than the economic questions raised by the first speaker, but it's a distraction from that issue—a red herring.

Perhaps the most blatant kind of red herring is evasive: used especially by politicians, this is the refusal to answer a direct question by changing the subject. Examples are almost too numerous to cite; to some degree, no politician ever answers difficult questions straightforwardly (there's an old axiom in politics, put nicely by Robert McNamara: "Never answer the question that is asked of you. Answer the question that you wish had been asked of you.").

A particularly egregious example of this occurred in 2009 on CNN's *Larry King Live*. Michele Bachmann, Republican Congresswoman from Minnesota, was the guest. The topic was "birtherism," the (false) belief among some that Barack Obama was not born in America and was therefore not constitutionally eligible for the presidency. After playing a clip of Senator Lindsey Graham (R, South Carolina) denouncing the myth and those who spread it, King asked Bachmann whether she agreed with Senator Graham. She responded thus:

> "You know, it's so interesting, this whole birther issue hasn't even been one that's ever been brought up to me by my constituents. They continually ask me, where's the jobs? That's what they want to know, where are the jobs?"

Bachmann doesn't want to respond directly to the question. If she outright declares that the "birthers" are right, she looks crazy for endorsing a clearly false belief. But if she denounces them, she alienates a lot of her potential voters who believe the falsehood. Tough bind. So she blatantly, and rather desperately, tries to change the subject. Jobs! Let's talk about those instead.

There is a special variant of the Red Herring fallacy that bears recognition: the smokescreen. Smokescreen strategies attempt to distract and divert the audience's attention away from the topic or inquiry at hand by way of a barrage of information that appears relevant to the topic. To the nefarious author, this has the added advantage of making them look thoughtful and well-informed. They seek to overwhelm their audience with important information that the audience is not ready to process. Along the way the discussion is tilted in a direction more favorable to the nefarious author.

Again, politicians provide some of the best examples. This is often because many of them are former lawyers who are well-versed in a complex legal system that touts highly technical language. Consider the following example:

> In a legal debate over proposed legislation to address police reform and accountability, a politician is asked about their position. Instead of directly addressing the question, the politician launches into a lengthy discourse about the intricacies of constitutional law, citing obscure legal precedents related to federalism and states' rights, and the historic difficulty of civilian oversight boards with subpoena powers. They employ convoluted legal jargon and complex terminology, effectively creating a smokescreen of legal complexity to obfuscate the straightforward issue of police accountability.

In this scenario, the politician uses a smokescreen fallacy by inundating the discussion with dense legal language, making it difficult for the audience to follow and diverting attention away from the core issue of police reform. This can be a delay tactic, or it can be used to soften an unpopular decision to do nothing and keep the status quo.

SET II

Appeals to Unwarranted Assumption

The fallacies in this group all share a common strategy in which the author of the argument makes use of a premise that they are unauthorized to use. We call these claims "unwarranted" to highlight the nature of the error.

Before we go further, we should first clarify the role of assumptions. Recall that an argument's basic structure is that of a set of statements, some of which are intended to support another.

> This basic structure reveals that in its most raw form, the premises of an argument are put forth unsupported. Their job is to support the truth of the conclusion, but in the first instance they themselves have no direct support in the argument.

One way to think of this is that the premises are all *assumed* to be true (by the author). If the audience challenges one or more of them, they will need to form a sub-argument to support the truth of the premise under attack, but initially this is not the intent. Make no mistake, *we count on assumptions in every argument*.

We also count on assumptions in daily life. This is unavoidable, and often there is nothing wrong with relying on an assumption.

The common quality of this group of fallacies is NOT that they involve assumptions. The common quality is that they involve assumptions that the author has no rightful claim to make. They help themselves to a claim that they are not entitled to make. Put differently, it is the "unwarranted" part of the assumption that is the source of the error, not the simple fact that an assumption was made.

Begging the Question

Begging the question occurs when one (either explicitly or implicitly) assumes the truth of the conclusion in one or more of the premises. Begging the question is thus a kind of circular reasoning.

Consider the following argument:

Capital punishment is justified for crimes such as rape and murder because it is quite legitimate and appropriate for the state to put to death someone who has committed such heinous and inhuman acts.

The premise indicator, "because," denotes the premise and (derivatively) the conclusion of this argument. In standard form, the argument is this:

1. It is legitimate and appropriate for the state to put to death someone who commits rape or murder.

Therefore, capital punishment is justified for crimes such as rape and murder.

You should notice something peculiar about this argument: the premise is essentially the same claim as the conclusion. The only difference is that the premise spells out what capital punishment means (the state putting criminals to death), whereas the conclusion just refers to capital punishment by name, and the premise uses terms like "legitimate" and "appropriate," whereas the conclusion uses a related term, "justified." But these differences don't add up to any real differences in meaning. Thus, the premise is essentially saying the same thing as the conclusion.

This is a serious failing of an argument: we want our premise to provide a reason to accept the conclusion. But if the premise is the same claim as the conclusion, then it can't possibly provide a reason for accepting the conclusion!

One interesting feature of this fallacy is that *formally* there is nothing wrong with arguments of this form. Here is what I mean. Consider an argument that explicitly commits the fallacy of begging the question. For example,

1. Capital punishment is morally permissible
2. Therefore, capital punishment is morally permissible

Now, apply any method of assessing validity to this argument and you will see that it is valid. If we use the informal test (by trying to imagine that the premises are true while the conclusion is false), then the argument passes the test, since any time the premise is true, the conclusion will have to be true as well (since it is the exact same statement).

While these arguments are technically valid, they are still really bad arguments. Why? Because the point of giving an argument in the first place is to provide some reason for thinking the conclusion is true for those who don't already accept the conclusion. But if one doesn't already accept the conclusion, then simply restating the conclusion in a different way isn't going to convince them. Rather, a good argument will provide some reason for accepting the conclusion that is sufficiently independent of the conclusion itself. Begging the question utterly fails to do this, and this is why it counts as an informal fallacy. What is interesting about begging the question is that there is absolutely nothing wrong with the argument *formally*.

Whether or not an argument begs the question is not always an easy matter to sort out. As with all informal fallacies, detecting it requires a careful understanding of the meaning of the statements involved in the argument and what the author is doing with them. Here is an example of an argument where it is not as clear whether there is a fallacy of begging the question:

> Christian belief is warranted because, according to Christianity, there exists a being called "the Holy Spirit," which reliably guides Christians towards the truth regarding the central claims of Christianity.[2]

One might think that there is a kind of circularity (or begging the question) involved in this argument, since the argument appears to assume the truth of Christianity in justifying the claim that Christianity is true. But whether this argument really does beg the question often rests on what the author actually means in their statements. Consider the term "warranted" in the above argument. There are at least two common ways to interpret this: as "understandable" or as "justified." Let's consider each in turn.

If the author means something like "understandable," then the argument clearly does not beg the question. After all, if the author wants you to believe that "It is understandable for Christians to adhere to the core tenets of their faith," then it is helpful to learn (in the premise) that they take these beliefs to come from a divine source. Now the audience has reason to believe in the truth of what the conclusion asserts: These beliefs are not "crazy" or "superstitious," but rather part of a sensible worldview in which beliefs should only derive from reliable sources.

However, consider if the term "warranted" means "justified" to the author. The argument becomes circular. This is because the conclusion is "Christian beliefs are justified" and the support is "Christian beliefs say their beliefs are justified." Now the audience has no reason to accept the truth of the conclusion unless they already accept the substance of the conclusion.

2. This is a much simplified version of the view defended by Christian philosophers such as Alvin Plantinga. Plantinga defends (something like) this claim in: Plantinga, A. 2000. *Warranted Christian Belief*. Oxford, UK: Oxford University Press.

Inappropriate Authority

In a society like ours, we have to rely on authorities to get on in life. For example, the things I believe about electrons are not things that I have ever verified for myself. Rather, I have to rely on the testimony and authority of physicists to tell me what electrons are like. Likewise, when there is something wrong with my car, I have to rely on a mechanic (since I lack that expertise) to tell me what is wrong with it. Such is modern life. So there is nothing wrong with needing to rely on authority figures in certain fields (people with the relevant expertise in that field)—it is inescapable. The problem comes when we invoke someone *whose expertise is not relevant* to the issue for which we are invoking it. For example:

> Bob read that a group of doctors signed a petition to prohibit abortions, claiming that abortions are morally wrong. Bob cites the fact that these doctors are *against* abortion, "See, even the people who know how it's done think it is wrong. Who else better to say? Abortion must be morally wrong."

Bob has committed the Appeal to Authority fallacy. The problem is that doctors are not authorities on what is morally right or wrong. Even if they are authorities on how the body works and how to perform certain procedures (such as abortion), it doesn't follow that they are authorities on whether or not these procedures should be performed—the ethical status of these procedures. It would be just as much an Appeal to Inappropriate Authority fallacy if Melissa were to argue that since some other group of doctors *supported* abortion, that shows that it must be morally acceptable. In either case (for or against), doctors are not authorities on moral issues, so their opinions on moral issues like abortion are irrelevant.

In general, an Appeal to Inappropriate Authority fallacy occurs when an author takes what an individual says as evidence for some claim, when that individual has no particular expertise in the relevant domain. This can be innocently done, especially when the source cited does have expertise in some other unrelated domain.

This is a favorite technique of advertisers. We've all seen celebrity endorsements of various products. Sometimes the celebrities *are* appropriate authorities. For example, there was a Buick commercial from 2012 featuring Shaquille O'Neal, the Hall of Fame basketball player, testifying to the roominess of the car's interior (despite its compact size). Shaq, a very, very large man, is an appropriate authority on the roominess of cars! But when Tiger Woods was the spokesperson for Buicks a few years earlier, it wasn't at all clear that he had any expertise to offer about their merits relative to other cars. Woods was an inappropriate authority; those ads committed the fallacy.

Usually, the inappropriateness of the authority being appealed to is obvious. But sometimes it isn't. A particularly subtle example is AstraZeneca's hiring of Dr. Phil McGraw in 2016 as a spokesperson for their diabetes outreach campaign. AstraZeneca is a drug manufacturing company. They make a diabetes drug called Bydureon. The aim of the outreach campaign, ostensibly, is to increase awareness among the public about diabetes, but of course the real aim is to sell more Bydureon. A celebrity like Dr. Phil can help. Q: Is he an appropriate authority? That's a hard question to answer. It's true that Dr. Phil had suffered from diabetes himself for 25 years, and that he personally takes the medication. So that's a mark in his favor, authority-wise. But is that enough? (HINT: No. It is not enough. We'll talk about how feeble Phil's sort of anecdotal evidence is in supporting claims about a drug's effectiveness when we discuss the hasty generalization fallacy. Suffice it to say, one person's positive experience doesn't prove that the drug is effective.)

You might be tempted to say,

> "But Dr. Phil isn't *just* a person who suffers from diabetes; he's a doctor! It's right there in his name. Surely that makes him an appropriate authority on the question of drug effectiveness."

This may sound appropriate, but Phil McGraw is not a *medical* doctor; he's a PhD (the "D" in PhD stands for "doctor of"). He has a doctorate in psychology. He's not even a licensed psychiatrist; he cannot legally prescribe medication. He has no relevant *professional expertise* about drugs and their effectiveness (especially with regard to diabetes). He is not an appropriate authority in this case. He looks like one, though, which makes this a very sneaky, but effective, advertising campaign.

False Dilemma / False Dichotomy

Suppose I were to argue as follows:

1. Raising taxes on the wealthy will either hurt the economy or help it.
2. But it won't help the economy.

Therefore, it will hurt the economy.

This argument contains a fallacy called a False Dilemma or "false dichotomy." A false dichotomy is simply a disjunction that does not exhaust all of the possible options. In this case, the problematic disjunction is the first premise: either raising the taxes on the wealthy will hurt the economy or it will help it. But these aren't the only options. Another option is that raising taxes on the wealthy will have no effect on the economy.

If we fail to see that the nefarious author has presented us with a false choice, we may evaluate the argument on its formal merits. If so, we'll note that it is a formally valid argument. However, since the first premise presents two options as if they were the only two options, when in fact they aren't, the first premise is false and the argument fails. This failure is due to its content, not its form. In this case, there is something more than merely a false premise. There is an active strategy to hide and/or limit the options from the audience. Additionally, the nefarious author's strategy will often disproportionately characterize one option in a negative light. The intent is to push the audience into the option that constitutes the nefarious author's conclusion. You feel like you are backed into a corner and have no other choice than to accept the author's conclusion.

Consider the following example. In a speech made on April 5, 2004, President Bush made the following remarks about the causes of the Iraq war:

> Saddam Hussein once again defied the demands of the world. And so, I had a choice: Do I take the word of a madman, do I trust a person who had used weapons of mass destruction on his own people, plus people in the neighborhood, or do I take the steps necessary to defend the country? Given that choice, I will defend America every time.

The False Dilemma here is the claim that: Either I trust the word of a madman or I defend America (by going to war against Saddam Hussein's regime). The problem is that these aren't the only options. Other options include ongoing diplomacy and economic sanctions—options which the holder of the highest political office in the land would have been *clearly* aware were on the table. Thus, even if it were true that Bush shouldn't have trusted the word of Hussein, it doesn't follow that the only other option was going to war against Hussein's regime. (Furthermore, it isn't clear why this was needed to defend America.) That is a false dichotomy.

Care should be taken when assessing an argument as a False Dilemma. The error of this fallacy lies in the falseness of the proposed "choice" between the two options. This is a strategic move made by a nefarious author. However, if the world is such that there really are only two options, we cannot lay blame on the world for limiting our choices to two. Consider the following:

> Shortly after the Titanic hits the iceberg, Leo tells Kate, "There's one last spot left. Get in the lifeboat without me or you're going to die!"

Also:

> Your friend is excited and tells you that her family has an extra ticket for you to go with them on their family vacation in Rome! You call your mom to tell her the news, and she tells you that your family is going to Hawaii…that same week.

To date, nobody has figured out how to be in two places at once. Similarly, no human body can survive for long in near-freezing water. Welcome to reality. Pouting about it and accusing your mom or Leo that they are committing a false dichotomy is not right. The world really is limiting your options. You *really are* faced with a *true* dilemma in both scenarios.

We only accuse an author of a False Dilemma when the two key hallmarks of a true dilemma are not met. *True* dilemmas can be identified by these two features:

a. The unavoidable constraint of choices (typically to two)
b. The equitable value of each option (typically both are roughly as bad as one another, but as the vacation example shows, both options can be nearly equally good as well)

Loaded Questions

Loaded questions are questions that presume the truth of some claim. Asking these can be an effective debating technique, a way of sneaking a controversial claim into the discussion without having outright asserted it.

The classic example of a loaded question is "Have you stopped beating your wife?" Notice that this is a yes-or-no question, and no matter which answer one gives, one admits to beating his wife. If the answer is "no," then the person continues to beat his wife; if the answer is "yes," then he admits to beating his wife in the past. Either way, he's a wife-beater. The question itself presumes the truth of this claim; that's what makes it "loaded."

Strategic deployment of loaded yes-or-no questions can be an extremely effective debating technique. If a nefarious author catches their opponent off-guard, they will struggle to respond to the question, since a simple "yes" or "no" commits them to the truth of the tacit presumption. This can make their opponent look evasive, shifty. And as they struggle to come up with a response, the nefarious author can pounce on them: "It's a simple question. Yes or no? Why won't you answer the question?"

Remember that we study the informal fallacies because they work. A loaded question strategy is a great way to *appear* to be winning a debate, even if you don't have a good argument. Imagine the following dialogue:

> Liberal TV Host: "Are you or are you not in favor of the president's plan to force wealthy business owners to pay their fair share in taxes to protect the vulnerable and aid this nation's underprivileged?"
>
> Conservative Guest: "Well, I don't agree with the way you've laid out the question. As a matter of fact…"
>
> Host: "It's a simple question. Should business owners pay their fair share: yes or no?"
>
> Guest: "You're implying that the president's plan would correct some injustice. But corporate taxes are already very…"
>
> Host: "Stop avoiding the question! It's a simple yes or no!"

The host might come off looking like the winner of the debate, with his opponent appearing evasive, uncooperative, and inarticulate.

Another use of loaded questions is the particularly sneaky political practice of "push polling." In a normal opinion poll, you call people up to try to discover what their views are about the issues. In a push poll, you call people up pretending to be conducting a normal opinion poll, pretending only to be interested in discovering their views, but with *a different intention* entirely: you don't want to know what their views are; *you want to shape their views*, to convince them of something. And you use loaded questions to do it.

A famous example of this occurred during the Republican presidential primary in 2000. George W. Bush was the front-runner, but was facing a surprisingly strong challenge from the upstart John McCain. After McCain won the New Hampshire primary, he had a lot of momentum. The next state to vote was South Carolina; it was very important for the Bush campaign to defeat McCain there and reclaim the momentum.

So a push poll was designed to spread negative feelings about McCain—by implanting false beliefs among the voting public. "Pollsters" called voters and asked, "Would you be more or less likely to vote for John McCain for president if you knew he had fathered an illegitimate black child?" The aim, of course, is for voters to come to believe that McCain fathered an illegitimate black child. But he did no such thing. He and his wife adopted a daughter, Bridget, from Bangladesh. Reference to this fact was not part of the poll. Instead, the poll question introduced negative associations with McCain as having a child out of wedlock from an interracial affair (both of which are severely frowned upon among the poll's intended audience).

A final note on loaded questions: there's a minimal sense in which every question is loaded. The social practice of asking questions is governed by implicit norms. One of these is that it's only appropriate to ask a question when there's some doubt about the answer. So every question carries with it the presumption that this norm is being adhered to: that it's a reasonable question to ask, that the answer is not certain. A nefarious author can exploit this fact, again to plant beliefs in listeners' minds that they otherwise wouldn't hold. For example, in a particularly shameful bit of alarmist journalism, the cover of the July 1, 2016, issue of *Newsweek* asks the question, "Can ISIS Take Down Washington?" The cover is an alarming, eye-catching shade of yellow, and shows four missiles converging on the Capitol dome. The simple answer to the question, though, is "no, of course not." There is no evidence that ISIS has the capacity to destroy the nation's capital. But the very asking of the question presumes that it's a reasonable thing to wonder about, that there might be a reason to think that the answer is "yes." The goal is to scare readers (*and sell magazines*) by getting them to believe there might be such a threat.

Gaslighting

There is a form of psychological abuse known as gaslighting. The name of this dysfunctional dynamic comes from the 1944 film *Gaslight*, which graphically depicts a husband who attempts to gain control over his wife by manipulating her self-perception and self-confidence.

Generally, the strategy is to engage in a sustained effort to undermine a person's trust in themselves. The goal is to gain power over the other person by gradually eroding their confidence in their own ability to evaluate their own perceptions and judge their reality for themselves.

When successful, the audience's self-doubt makes them susceptible to accepting the authority of the nefarious author who now stands as the only "reliable source" of truth. The nefarious author's conclusions are readily accepted because the audience is no longer able to accept their own judgments.

We should note that real-world instances of this occur over long periods of time and through a range of tactics, many of which do not directly involve arguments. However, there are common argumentative tactics that closely map to the broader gaslighting strategy. Such a nefarious author will frequently pose arguments of the following form:

1. You say you saw (hear, suspect, or think) that X is the case.
2. You are clearly not well; something is wrong with you to imagine that X could be the case.

Therefore, X is not the case, and you should consider what caused you to make such a mistake.

Note that two elements of this strategy are in play:

a. Denying that the audience's perception of reality is accurate
b. Placing fault for this on the audience

There are number of ways each element can be attempted. The most obvious way we see element (a) play out is simple denial ("It didn't happen"). In slightly more subtle ways, a nefarious author can also feign ignorance ("I don't know anything about that") or offer counterfactual accounts ("That's not how it happened"). Of course, denying the truth of a claim does not automatically mean that an author is initiating a gaslighting strategy. People often legitimately disagree about whether or not something is true. In a gaslighting strategy, the denial is absolute and tied to those who make the denied claim. Often the nefarious author will simply deny the claim with no attempt to offer any support, even when it is directly requested. The denial is persistent and linked to the audience's perception. They are not simply denying the truth of a claim, they are denying the ability to affirm it.

Ironically, in many cases, the more excessively (a) is carried out, the more it plays into element (b) by seamlessly blending the two together. For example:

> "*What are* you *talking about? It never happened!* You *keep going on about it, but that wasn't me!* You *are imagining things.*"

Notice how this author has now slipped in references to the audience—*they* are the source of the claims, and those claims are clearly false. For the nefarious author, the audience needs to be reminded that it is *their account* of reality that is deeply flawed, so much so that only a defective person could believe it. "*You know you sound crazy, don't you? I mean, just listen to what you're saying. People are going to start talking if you keep going on like this*." The nefarious author's goal is to establish fault with the audience—they are to blame for thinking (anything that the author does not want them to think).

In a healthy argument in which people disagree, it is common to deny another person's claims. However, the denial of these claims is based on other pieces of evidence. Each party tries to bring claims about the subject of their disagreement into their respective arguments (e.g., "I disagree that you shouldn't have to pay for half of the repair bill. You use the car just as much as I do, and it needs to get fixed."). Gaslighting tactics do not focus on the subject of the disagreement. Instead, they rely on psychological assaults on the other side's very capacity to make judgments (a kind of specialized ad hominem with a more insidious objective than simply discrediting a single conclusion):

> "*I never so much as looked at that woman. You know, this is really becoming pathological already; you have to stop, we can't go on with you acting this way*."

Healthy disagreements also involve listening to one another and considering the merits of the other party's claims. Gaslighting tactics are particularly keen to minimize, ignore, or deflect the claims of the other party.

> "*It's not that big a deal*."
>
> "*You always overreact like this.*"
>
> "*Oh sure, but what about your behavior?*"

Phrases like these ensure that the nefarious author does not need to listen to a counterclaim. Quite the contrary, the audience needs to now reconsider why they thought the claim merited voicing in the first place.

Before leaving this fallacy, we should note that classic gaslighting efforts are interpersonal. They occur from person to person. However, a social version of this fallacy can occur when groups of people are methodically manipulated into no longer trusting their own ability to engage with and judge public discourse (i.e., news, media, social policy, institutional decisions).

The nefarious authors in these cases are typically politicians and public speakers (talk show hosts, social media influencers, politicians, etc.). They can engage in extended campaigns to slowly erode the public's confidence in social institutions (our courts/legal system, political parties/processes, school systems/public education, and even general scientific practices). Just as with a person-to-person case of gaslighting, when there is no way for the audience to trust their own ability to perceive reality and judge for themselves, the audience will turn to the nefarious author for guidance. Large groups of people will turn to the very people who insist they cannot trust their own eyes. They will believe even the most bald-faced lies, because their only remaining "reliable source" is the nefarious author.

SET III

Weak Induction

As the name suggests, what these fallacies have in common is that they are poor inductive arguments. That is, our technical evaluation of them would say they are weak arguments. Recall that inductive arguments attempt to provide premises that make their conclusions more probable. We evaluate them according to how probable their conclusions are in light of their premises: the more probable the conclusion (given the premises), the stronger the argument; the less probable, the weaker. The fallacies of weak induction are arguments whose premises do not make their conclusions very probable—but that are nevertheless often successful in convincing people of their conclusions.

Hasty Generalization

Inductive arguments involve an inference from particular premises to a general conclusion; this is generalization. For example, if you observe every morning that the sun rises in the east, and conclude on that basis that, in general, the sun always rises in the east, this is a generalization. And it's a good one! With all those sunrise observations as premises, your conclusion that the sun always rises in the east has a lot of support; that's a strong inductive argument.

One commits the hasty generalization fallacy when one makes this kind of inference based on an insufficient number of premises, when one is too quick—hasty—in inferring the general conclusion. This makes for a weak inductive argument; the evidence provided in the premises does not provide much assurance that the general claim or claims made in the conclusion are true.

People who deny scientific claims often make this mistake—particularly when they themselves are not scientists and are not adhering to basic standards of scientific reasoning. For example, many people will deny that global warming is a genuine phenomenon and, often enough, attempt to support their denial by committing this fallacy.

> In February of 2015, the weather was unusually cold in Washington, DC. Senator James Inhofe of Oklahoma famously took to the Senate floor wielding a snowball. "In case we have forgotten, because we keep hearing that 2014 has been the warmest year on record, I ask the chair, 'You know what this is?' It's a snowball, from outside here. So it's very, very cold out. Very unseasonable."

Senator Inhofe commits the hasty generalization fallacy. He's trying to establish a general conclusion—that 2014 wasn't the warmest year on record, or that global warming isn't really happening (he's on the record that he considers it a "hoax"). But the evidence he presents is insufficient to support such a claim. His evidence is an unseasonable coldness in a single place on the planet, on a single day. We can't derive from that any conclusions about what's happening, temperature-wise, on the entire planet over a long period of time.

Many people are susceptible to hasty generalizations in their everyday lives. When we rely on limited anecdotal evidence to make decisions, we commit the fallacy. Suppose you're thinking of buying a new car, and you're considering a Subaru. Your neighbor has a Subaru. So what do you do? You ask your neighbor how he likes his Subaru. He tells you it runs great, hasn't given him any trouble. You then, fallaciously, conclude that Subarus must be terrific cars. But one person's testimony isn't enough to justify that conclusion; you'd need to look at many, many more drivers' experiences to reach such a conclusion.

A particularly pernicious instance of the Hasty Generalization fallacy is the development of negative stereotypes. People often make general claims about religious or racial groups, ethnicities and nationalities, based on very little experience with them. If you once got mugged by a Puerto Rican, that's not a good reason to think that, in general, Puerto Ricans are crooks. If a waiter at a restaurant in Paris was snooty, that's no reason to think that French people are stuck up. And yet we see this sort of faulty reasoning all the time.

Cherry Picking / Confirmation Bias

The Cherry Picking or Confirmation Bias fallacy occurs when an author has a particular way of using evidence to demonstrate support for their conclusion. Normally, we would hope that someone would take in all of the available evidence about a particular subject, weigh its relative credibility, and then come to a conclusion. The Cherry Picking fallacy happens when someone *already knows* which conclusion they'd like to prove and *then selects evidence* that supports that conclusion. Instead of looking at the evidence and letting it determine what conclusion we draw, Confirmation Bias leads us to use a fixed conclusion to determine how we "pick" the evidence we will pay attention to and offer up to others. It's backwards.

There's a humorous story of this backwards mental strategy that goes as follows:

> Once there was a man in Texas who shot at his barn door with his rifle. When he had unloaded ten rounds, he walked up to the door, found a cluster of bullet holes that were particularly tightly clustered, and painted a bullseye around them. He then bragged to his friends (and warned his enemies) that here was the evidence that he was a fantastic sharpshooter!

If you only weigh the evidence that supports the conclusion you like (or in the story, if you only draw the target around the bullet holes that look good), then you'd be disregarding other evidence for no other reason than that it got in the way of your concluding what you wanted to conclude.

Here's another example.

> You have a friend who thinks vaccines are unsafe. You ask why, and they present you with lots of evidence they found. They present this evidence as though there is absolutely nothing to say on the other side of this complex matter. Sure, they gathered evidence on their side, like anecdotes about apparent vaccine injuries, the medical professionals who make claims about vaccines being dangerous, and the apparent empirical evidence that connects vaccines and illnesses of various kinds. However, your friend

shows additional signs that they are completely unaware of any other view. When we mention some of the reasons people may have a different view, they are prone to ignore, discount, or explain away all of the evidence that seems to show that vaccines aren't significantly connected with illness or injury.

The core error here is that your friend has let his desired conclusion determine which evidence he was willing to take into account. We may also find that he did not treat evidence equitably. You and your friend may have engaged in a profitable discussion, disagreeing and debating as the evidence in the case allowed. Such is the nature of evidence when it is responsibly evaluated. But this did not happen, because your friend was not open to making conclusions *on the basis of* the evidence; they were only open to evidence *when it backed up their pre-existing* conclusions. A classic case of confirmation bias.

Appeal to Ignorance

This is a particularly egregious fallacy. In essence, it's an inference from premises that assert a lack of knowledge about some topic to a definite conclusion about that topic. In other words: *We don't know; therefore, we know!*

Of course, put that baldly, it's plainly absurd; actual instances are more subtle. The fallacy comes in a variety of closely related forms. It will be helpful to state them in bald/absurd schematic fashion first, then elucidate with more subtle real-life examples.

The first form can be put like this:

> Nobody knows how to explain phenomenon X.
> ∴ My theory about X is true.

That sounds silly, but consider an example: there is no shortage of "documentary" programs on cable TV about space aliens. These programs suggest that extraterrestrials built the pyramids or something like that. How do they get you to believe that crazy theory? Simple: they use a strategy that gets the audience to fill in the blanks. But you may ask, how do you do that? Easy: By creating mystery! By pointing to facts that nobody can explain. Here's what it often looks like:

> The Great Pyramid at Giza is aligned (almost) exactly with the magnetic north pole! On the day of the summer solstice, the sun sets exactly between two of the pyramids! The height of the Great Pyramid is (almost) exactly one one-millionth the distance from the Earth to the Sun! How could the ancient Egyptians have such sophisticated astronomical and geometrical knowledge? Why did the Egyptians, careful record-keepers in (most) other respects, (apparently) not keep detailed records of the construction of the pyramids? Nobody knows. Conclusion: aliens built the pyramids.

In other words, there are all sorts of surprising facts about the pyramids, and nobody knows how to explain them. From these premises, which establish only our ignorance, we're encouraged to conclude that we know something: aliens built the pyramids. That's quite a leap—too much of a leap.

Another form this fallacy takes can be put crudely thus:

> Nobody can PROVE that I'm wrong.
> ∴ I'm right.

The word "prove" is in all-caps because stressing it is the key to this nefarious strategy: the standard of proof is set impossibly high, so that almost no amount of evidence would constitute a refutation of the conclusion.

Yet another version of the Argument from Ignorance goes like this:

I can't imagine/understand how X could be true.

∴ X is false.

Of course, lack of imagination on the part of an individual isn't evidence for or against a proposition, but people often argue this way. The world is a wondrous place; its facts care little for your powers of understanding or imagination. If some particular claim is true, we must strive for direct evidence that supports it, not simply *appeal to our failings* as human beings.

A final form of the Argument from Ignorance can be put crudely thus:

No evidence has been found that X is true.

∴ X is false.

You may have heard the slogan, "Absence of evidence is not evidence of absence." This is an attempt to sum up this version of the fallacy. But it's not quite right. What it should say is that absence of evidence is not always definitive evidence of absence. An example will help illustrate the idea.

> During the 2016 presidential campaign, a reporter (David Fahrentold) took to Twitter to announce that despite having "spent weeks looking for proof that [Donald Trump] really does give millions of his own [money] to charity…" he could only find one donation, to the NYC Police Athletic League. Trump has claimed to have given millions of dollars to charities over the years.

Does this reporter's failure to find evidence of such giving prove that Trump's claims about his charitable donations are false? No. To rely only on this reporter's testimony to draw such a conclusion would be to commit an Appeal to Ignorance fallacy.

However, the failure to uncover evidence of charitable giving does provide some reason to suspect Trump's claims may be false. How much of a reason *depends on* the reporter's methods and credibility, among other things. But sometimes a lack of evidence can provide strong support for a negative conclusion. This is an inductive argument; it can be weak or strong. For example:

> Despite multiple claims over many years (centuries, if some sources can be believed), no evidence has been found that there's a sea monster living in Loch Ness in Scotland. Given the size of the body of water, and the extensiveness of the searches, this is pretty good evidence that there's no such creature.

This would be a strong inductive argument to that conclusion. Conversely, to claim otherwise—that there is such a monster, despite the near-complete lack of evidence—would be to commit the version of the fallacy where one argues "You can't PROVE I'm wrong; therefore, I'm right." Here, the nefarious author appeals to a guaranteed lack of evidence (because the standard of proof is unreasonably high).

One final note on this fallacy: it's common for people to mislabel certain bad arguments as arguments from ignorance; namely, arguments made by people who obviously don't know what the heck they're talking about. People who are confused or ignorant about the subject on which they're offering an opinion are liable to make bad arguments, but the fact of their ignorance is not enough to label those arguments as instances of the fallacy.

After all, in these cases the author of those arguments is less a "nefarious" character and more of a "witless" character. Put differently, the witless author is not using a *strategy* of not-knowing—they simply don't know what they are talking about, and it can be painfully obvious that they are not aware of this. These are *not* Appeals to Ignorance.

Conspiracy Theory (a.k.a. Canceling Hypothesis)[3]

An important variation on the Appeal to Ignorance is the Conspiracy Theory (more generally, the Canceling Hypothesis). The argument defends one hypothesis (i.e., claim #1) by proposing a second hypothesis (claim #2) to explain the *lack* of evidence in support of the first hypothesis. That is, the second hypothesis cancels or undermines the predictions made by the first hypothesis.

> The general strategy is to present a conclusion and then support or defend it from attack by using another claim to explain why *there is no support* for the conclusion.

In this sense, the strategy is a special application of Appeal to Ignorance. The nefarious author has no actual support for their conclusion. The key feature of this variation is that the author does not simply appeal to a general lack of knowledge and leave it at that; they try to *explain away their ignorance* by appeal to a second claim (a second hypothesis, which itself is not supported).

Conspiracy theories are a particularly important example of this fallacy. The proposed conspiracy (i.e., the second claim/hypothesis) generally involves a conscious effort to destroy or cover up the evidence that would otherwise have supported a given hypothesis. However, the term "conspiracy theory" is generally used only when there has been a conscious effort to hide evidence, and not all examples of Canceling Hypotheses involve such conscious effort.

Three examples:

> "The earth was created just six thousand years ago. *The reason atomic dating and erosion patterns suggest a much older Earth is that* God created the earth in an 'aged' condition."
>
> "Based on a theory that brain size would have some correlation with personality, Paul Broca predicted that criminals would have smaller average brain size than non-criminals. Weighing of brains failed to support this hypothesis, *but Broca noted that the criminals in his study had* died quickly of violent deaths, while the non-criminals had died of long illnesses, during which their brains had probably diminished in size."
>
> "Strange lights and unusual military activity in 1947 around Roswell, New Mexico, were due to a

3. Bruce Thompson claims: I first became aware that this was an important fallacy in scientific reasoning from reading Stephen Jay Gould's The Mismeasure of Man, New York: Norton, 1981. I coined the ugly term "canceling hypotheses." The term "conspiracy theory" is catchier, and has greater popular currency (see, for example, the movie of that title starring Mel Gibson); but it must be remembered that not all instances of this fallacy involve a conspiracy.

crash landing (or even possibly the shooting down) of an alien spacecraft. The government *denies that this is so and is covering up the evidence* so the public will not panic."

The first two examples illustrate the proposal of a non-intentional Canceling Hypothesis, while the third invokes an intentional Conspiracy Theory. In all examples, note how the elements in italics signal the introduction of the second claim that explains the lack of evidence for the first claim. Of course, this first claim is the nefarious author's conclusion, so they can assert it as true *even though* there is no evidence for it.

In a full-blown Conspiracy Theory, the strategy is to assert that the conclusion must be true *especially because* there is no evidence. Any potential evidence (so the second claim goes) was eliminated or removed by the members of the alleged conspiracy—the harder it is to find such evidence, the deeper the conspiracy goes. For example:

"We will never have a fair election in this country; the CIA has rigged the election by hacking into voting machine software—*oh, but you'll never see it*; they're too good at this. And these members of the Deep State are in every part of the government covering each other's backs."

Sometimes, of course, there are factors that hide evidence. Much evidence about ancient civilizations has been destroyed by grave robbers, for example. The fossil record is notoriously incomplete, due in part to erosion and other geological forces. Sometimes there even *are* intentional cover-ups, as when Nixon's staff intentionally hid evidence of the Watergate break-in. However, we discover what factors are responsible for hiding evidence by *looking at other evidence*. We expect that evidence will eventually come to light—even in the case of an intentional cover-up—provided that we continue looking for evidence.

Conspiracy Theory (Canceling Hypotheses) attempts to shut down further inquiry, rather than keep the investigation going. It is an attempt to defend the original hypothesis *against lack of evidence*; it is not an invitation to look for evidence supporting the second (canceling) hypothesis. Indeed, in a really elaborate conspiracy theory, it may be necessary to propose a third hypothesis to explain why we can find no evidence of a cover-up! "*Yeah, you say they have looked into it and found no evidence, but that just proves how deep the cover-up goes!*" By shutting down the investigation, it effectively makes the hypothesis untestable.

Faulty Causality (a.k.a. False Cause)

The most common form of the Faulty Causality fallacy is captured in the admonition:

Correlation does not equal causation.

The mere fact that some event correlates roughly in time with some other event does not mean we have good reason to think it caused the other event. There are several classic ways in which people make mistakes in reasoning about causality, but we will begin here.

This first variant of this fallacy occasionally goes by a Latin term, *post hoc ergo propter hoc*, usually shortening it to "post hoc." The whole phrase translates to "After this, therefore because of this," which is a pretty good summation of the pattern of reasoning involved.

Crudely, the schematic looks like this:

1. X occurred before Y.

∴ X caused Y.

This is not a good inductive argument. That one event occurred before another gives you some reason to believe it might be the cause—after all, X can't cause Y if it happened after Y did—but not nearly enough to conclude that it is the cause.

Many people reason poorly about "the power of prayer" and other such miracles in this way. Note that the poor reasoning is specifically about the "power" part of these claims. Here, "power" means "causal efficacy." *So, the concern is not with the nature of faith or religious belief in general—the concern is with how claims of causal efficacy are used in arguments*. For example,

> I know prayer works. You remember Matthew; he was diagnosed with cancer last year. Well, we all prayed for his recovery week after week. Guess what! He just saw his doctor and was declared cancer-free!

Note that the author did not simply declare their religious beliefs. The author has made an argument. This argument makes use of the classic Faulty Causality (post hoc) fallacy. We can readily identify this on two fronts:

> First, the basic structure of the claim "we did this, and it led to that" was used to identify a causal claim.
>
> Second, the tacit acknowledgement that there were *other factors* at play: Matthew *was* seeing a doctor. However, these factors were not considered as causally significant.

The second feature of the argument reveals another classic variant of the Faulty Causality fallacy: failing to recognize/acknowledge other relevant causal factors. Making accurate and reliable causal claims is very difficult. Great care must be taken to ensure we don't miss contributing factors. You may remember your science experiments in school—many of the procedures you used were intended to carefully account for all the variables that could impact the results. In the actual world, causality is messy. There's a great deal that feeds into events.

Consider our example again. Did our author above know what therapies Matthew was using all this time? Does the author consider the changes he made in his medical provider, the improved and more accurate diagnosis he got last fall, the experimental medication that Matthew tried? Does the author even consider Matthew's general health or the general probability of recovery from the type of cancer afflicting him? A person's overall health is highly complex, and many variables factor into it. Singling out one thing (that happened in the general timeframe under discussion) and claiming that *this* caused improvement in a person's health grossly ignores the bulk of what needs to be considered when making such causal claims.

Nowhere is this fallacy more in evidence than in our evaluation of the performance of presidents of the United States. Everything that happens during or immediately after their administrations tends to be pinned on them. But presidents aren't all-powerful; they don't cause everything that happens during their presidencies.

On July 9th, 2016, a short piece appeared in the *Washington Post* with the headline "Police are safer under

Obama than they have been in decades." What does a president have to do with the safety of cops? Very little, especially compared to *other factors* like poverty, crime rates, policing practices, rates of gun ownership, etc.

Another example: in October 2015, *U.S. News & World Report* published an article asking (and purporting to answer) the question, "Which Presidents Have Been Best for the Economy?" It had charts listing GDP growth during each administration since Eisenhower. But while presidents and their policies might have some effect on economic growth, their influence is certainly swamped by other factors. More importantly, the causal influence of different events in an economy are almost always extended over long periods of time. Nefarious authors rarely acknowledge that what happens in the economy *this year* is partially the result of the policies put into place by the prior administration.

Attempts to win political points frequently abuse this fallacy. At the 2016 Republican National Convention, Governors Scott Walker and Mike Pence—of Wisconsin and Indiana, respectively—both pointed to record-high employment in their states as vindication of their conservative, Republican policies. But some other states were also experiencing record-high employment at the time: California, Minnesota, New Hampshire, New York, Washington. Yet they were all controlled by Democrats. Maybe there's a separate cause for those strong jobs numbers in differently governed states? Possibly it has something to do with the improving economy and overall health of the job market in the whole country.

Slippery Slope

The Slippery Slope fallacy is a variant of the general Faulty Causality fallacy. This fallacy is committed when one event is said to lead to some other (usually disastrous) event via a chain of intermediary events. The strategy is to claim that the first step in this causal chain will result in an unstoppable sequence (like stepping out on a slippery surface—once you get started you cannot stop yourself), so the only responsible choice is to avoid that first step entirely. The nefarious author's conclusion is always "avoid that first step" (whatever that "step" may be).

If you have ever seen DirecTV's "get rid of cable" commercials, you will know exactly what I'm talking about. (If you don't know what I'm talking about, you should Google it right now and find out. They're quite funny.) Here is an example of a causal slippery slope fallacy (it is adapted from one of the DirecTV commercials):

> If you use cable, your cable will probably go on the fritz. If your cable is on the fritz, you will probably get frustrated. When you get frustrated, you will probably hit the table. When you hit the table, your young daughter will probably imitate you. When your daughter imitates you, she will probably get thrown out of school. When she gets thrown out of school, she will probably meet undesirables. When she meets undesirables, she will probably marry undesirables. When she marries undesirables, you will probably have a grandson with a dog collar. Therefore, if you use cable, you will probably have a grandson with a dog collar.

This example is silly and absurd, yes. But it illustrates the slippery slope fallacy. Slippery slope fallacies are always made up of a series of conditional statements that link the first event to the last event.

Two broad strategic variations are worth noting:

1. Dubious links in the chain
2. Dubious length of the chain

Both reflect different ways to commit this fallacy, and both require careful attention to how the author constructs and uses the links in the series of events.

First, in most cases of Faulty Causality the links in the causal chain are dubious. If you look at any one of them in isolation, there is often little reason to accept the probability of such causal claims. However, the nefarious author does not worry about that, because the whole point is to make it *look like* they are presenting a well thought out series of events. The author wants to present their claims like a series of dominos that are ready to fall.

Touch that first one and you will see the rest inevitably fall.

The presentation of the series itself is what they hope lends credibility to their argument. Of course, looking closely at that string of dominos reveals that they are playing with magic dominos—they would have to leap rather remarkably and with uncanny accuracy to make it to the next domino.

Second, sometimes you look at the individual links in a chain and (when considered in isolation) they seem probable. This is an even more compelling version of the fallacy, and we must have some knowledge of how to reason about probability in order to spot the error.

If we are not careful, we will accept the individual claims the nefarious author provides, and then accept their conclusion without realizing that it is based on the *cumulative probability* of all their claims *put together*.[4] This significantly reduces the probability that the first step will lead to the last terrible outcome. Yet the nefarious author hopes that you do not know this, and that you will instead think of the overall probability of the chain as equivalent to the individual (high) probabilities of each link in it.

Gambler's Fallacy

The gambler's fallacy occurs when one thinks that independent, random events can be *influenced* by each other. This is a common tendency in how people reason about causality. Thus, we regard this as another special variant of the Faulty Causality fallacy.

For example, suppose I have a coin and I have just flipped 4 heads in a row. Erik also has a coin (a regular coin, just like mine, nothing particularly suspect about either). He has flipped his coin 4 times and gotten tails. We are each taking bets that *the next* coin flipped is heads. Who should you bet flips the head?

---

4. Mathematically speaking, when you do this, you have to multiply the probabilities of each claim, making the result for the overall chain of events lower and lower with every added link.

Go ahead. Take a bit of time to consider this scenario and make your bet…think it through.

No, no, really, take a bit more time. *Don't cheat this example*. Re-read the example and think about who you would bet flips the head.

If you are inclined to say that you should place the bet with Erik since he has been flipping all tails and, since the coin is fair, the flips must even out soon, then you have committed the gambler's fallacy. The fact is, each flip is independent of the next, so the fact that I have just flipped 4 heads in a row does not increase or decrease my chances of flipping a head. Likewise for Erik.

It is true that as long as the coin is fair, then over a large number of flips we should expect that the proportion of heads to tails will be about 50/50. But there is no reason to expect that *a particular flip* will be more likely to be one or the other. Since the coin is fair, each flip has the same probability of being heads and the same probability of being tails—50%.

This can be a bit hard to wrap our minds around. We think: *Surely, the next flip will be* (insert opposite of what it has regularly been). We usually commit this fallacy on ourselves (TIP: don't go to Vegas and play the roulette wheel without understanding this clearly). If you are taking an exam and there are many true/false questions, we often look at the pattern of true and false answers and use that to consider the probability of the next question's answer. *The last five were all "true," so the next one is probably "false."* Not a good idea.

Precisely because we (normal people, like you and me) do this to ourselves and tend to struggle with this, a nefarious author can make use of this tendency. Clearly, if the nefarious author has insider information and knows the outcome of the next instance, they can leverage this against us by offering a bet in accordance with what we're inclined to fallaciously predict. Fortunately, this is rare (because few nefarious authors are actual evil geniuses who have this diabolical knowledge). What is more common is the use of this fallacy to cover one's tracks. Consider:

> Luis and Jane are married, and though they are struggling financially, they have managed to save some money for their kid's college. Luis has been watching the Mega Millions lottery, and for 16 weeks straight there has been no winner. He tells Jane, "There's got to be a winner soon! Let's take out the savings and put it on a ton of numbers." After the 17th week there are no winners, and Luis and Jane lose their entire savings. Luis defends himself: "Don't blame me! You knew the odds; we took our best bet and made a run at it. This thing is rigged, babe, there just *had to be* a winner."

The likelihood that a winner would be found this week does not change just because there has been no winner up until now (assuming the same number of tickets are sold this week as in previous weeks). Unfortunately, Luis played himself *and* his wife by falling victim to this fallacy while at the same time promoting it to his wife.

A Final Word on Social Phenomena

We should note two things before leaving our review of informal fallacies:

1. Nothing prevents someone from committing multiple informal fallacies in a single sitting; any nefarious author can layer fallacies on top of fallacies
2. Broad social maladies we witness today are a prime example of where you can find layered informal fallacies

Sociopolitical phenomena like racism, imperialism, cancel culture, and formally organized conspiracy theories are extremely complex in their underlying causes. However, as logicians we should recognize the way collections of frequently grouped fallacies are used to perpetuate these ills.

For example, the racism we encounter today uses a collection of ad hominem, hasty generalization, and cherry-picking fallacies to both further racist arguments as well as justify (or deny) racist conduct. Similarly, when we study actual conspiracy theories, we easily see that they are anchored in the fallacy of conspiracy theory, but they also tend to involve gaslighting, bandwagon appeals, faulty causality, and loaded questions.

3.

INTRODUCTION TO PRINCIPLES OF INDUCTIVE REASONING—ANALOGY AND CAUSALITY

Types of Inductive Logic

Our previous chapter helped us see that there are many ways in which informal errors can find their way into our reasoning. While this is very helpful, we should also consider positive ways in which our thinking can be sharpened. Put another way, knowing what *not to do* is good, but so too is knowing *what we should do*.

Valid reasoning is one thing we should do, and we will explore this in much greater detail later. However, valid reasoning is not the only kind of "good" reasoning. So, even when we are not striving for formally valid reasoning, the fates do not doom us to commit informal errors. We can still do a fine job thinking about things if we know some basic principles of inductive reasoning.

Back in Chapter 1, we made a distinction between deductive and inductive arguments. While deductive arguments attempt to provide premises that guarantee their conclusions, inductive arguments are less ambitious. They merely aim to provide premises that make the conclusion more probable. Because of this difference, it is inappropriate to evaluate deductive and inductive arguments by the exact same standards.

We do not use the terms "valid" and "invalid" when evaluating inductive arguments: technically, they're all invalid because their premises don't guarantee their conclusions, but that's not a full and fair evaluation, since inductive arguments don't even pretend to provide such a guarantee. Rather, we say of inductive arguments that they are strong or weak—the more probable the conclusion in light of the premises, the stronger the inductive argument; the less probable the conclusion, the weaker. These judgments can change in light of new information. Additional evidence may have the effect of making the conclusion more or less probable—of strengthening or weakening the argument.

The topic of this chapter will be inductive logic: we will be learning about the various types of inductive arguments and how to evaluate them. Inductive arguments are a rather motley bunch. They come in a wide variety of forms that can vary according to subject matter; they resist the uniform treatment we were able to provide for their deductive cousins. We will have to examine a wide variety of approaches—different inductive logics.

While all inductive arguments attempt to make their conclusions more probable, it is not always possible for us to make precise judgments about exactly how probable their conclusions are in light of their premises.

When that is the case, we will make relative judgments: this argument is stronger or weaker than that argument, though I can't say how much stronger or weaker, precisely. Sometimes, however, it will be possible to render precise judgments about the probability of conclusions, so it will be useful for us to acquire basic skills in calculating probabilities. For present purposes, this textbook will not delve into statistical analysis.

In this chapter, we will look at two very common types of inductive reasoning: arguments from analogy and inferences involving causation. The former are quite common in everyday life; the latter are the primary methods of scientific and medical research.

Each type of reasoning exhibits certain patterns, and we will look at the general forms of analogical and causal arguments; we want to develop the skill of recognizing how particular instances of reasoning fit these general patterns. We will also learn how these types of arguments are evaluated. Generally, for **analogy**, we will identify the criteria that we use to make relative judgments about strength and weakness. For **causal reasoning**, we will compare the various forms of inference to identify those most likely to produce reliable results, and we will examine some of the pitfalls peculiar to each that can lead to errors.

Arguments from Analogy

Analogical reasoning is ubiquitous in everyday life. We rely on analogies—similarities between present circumstances and those we've already experienced—to guide our actions. We use comparisons to familiar people, places, and things to guide our evaluations of novel ones. We criticize people's arguments based on their resemblance to obviously absurd lines of reasoning. In this section, we will look at the various uses of analogical reasoning. Along the way, we will identify a general pattern that all arguments from analogy follow and learn how to show that particular arguments fit the pattern. We will then turn to the evaluation of analogical arguments: we will identify six criteria that govern our judgments about the relative strength of these arguments. Finally, we will look at the use of analogies to refute other arguments.

The Form of Analogical Arguments

Perhaps the most common use of analogical reasoning is to predict how the future will unfold based on similarities to past experiences. Consider this simple example. When I first learned that the movie *The Wolf of Wall Street* was coming out, I predicted that I would like it. My reasoning went something like this:

> *The Wolf of Wall Street* is directed by Martin Scorsese, and it stars Leonardo DiCaprio. Those two have collaborated several times in the past, such as on *Gangs of New York*, *The Aviator*, *The Departed*, and *Shutter Island*. I liked each of those movies, so I predict that I will like *The Wolf of Wall Street*.

Notice, first, that this is an inductive argument. The conclusion, that I will like *The Wolf of Wall Street*, is not guaranteed by the premises; as a matter of fact, my prediction was wrong and I really didn't care for the film. But our real focus here is on the fact that the prediction was made based on an analogy between *The Wolf* *of Wall Street*, on the one hand, and all the other Scorsese/DiCaprio collaborations on the other. The general pattern was something like this:

> The new film is similar in important respects to the older ones; I liked all of those; so, I'll probably like the new one.

We can use this pattern of reasoning for more overtly persuasive purposes. Consider the following:

> Eating pork is immoral. Pigs are just as smart, cute, and playful as dogs and dolphins. Nobody would consider eating those animals. So why are pigs any different?

That passage is trying to convince people not to eat pork, and it does so on the basis of analogy: pigs are just like other animals we would never eat—dogs and dolphins.

Analogical arguments all share the same basic structure. We can lay out this form schematically as follows:

1. The things a1, a2, … and *C* all have the properties P1, P2, …
2. a1, a2, … all have property Q

So, *C* has Q

This is an abstract schema, and it's going to take some getting used to, but it represents the form of analogical reasoning succinctly and clearly. Arguments from analogy can be represented as having two premises and a conclusion.

The first premise establishes an analogy. The analogy is between some thing, marked "*C*" in the schema, and some number of other things, marked "a1," "a2," and so on in the schema. We can refer to these a-things as the "analogues" while the *C*-thing is referred to as the "target." Analogues are the things that are similar to the target. We say they are analogous to the target *C*. This schema is meant to cover every possible argument from analogy, so we do not specify a particular number of analogues; this is why we end the list with "…" to show there could be any number of them. There may be only one analogue; there may be a hundred. What's important is that the analogues are similar to the thing designated by "*C*."

What makes different things similar? They have stuff in common; they share properties. Those properties—the similarities between the analogues and *C*—are marked "P1," "P2," and so on in the diagram. Again, we don't specify a particular number of properties shared, so we use the "…" to represent this (the number of analogues and the number of properties can of course be different). Our schema is intentionally generic: every argument from analogy fits into the framework; there may be any number of properties involved in any particular argument. Anyway, the first premise establishes the analogy: *C* and the analogues are similar because they have various things in common—P1, P2, P3, ….

Notice that the target "*C*" is missing from the second premise. The second premise only concerns the analogues: it says that they have some property in common, designated "Q" to highlight the fact that it's not among the properties listed in the first premise. Q is a separate property; it's the very property that we are trying to say *C* has in the conclusion. This is why we call *C* the target: we are trying to pin Q onto *C*, for it is the target we aim to hang it on. The thinking is something like this:

C and the analogues are similar in so many ways (first premise)

The analogues have this additional thing in common (Q in the second premise)

So, *C* is probably like that, too (our conclusion: *C* has Q).

It will be helpful to apply these abstract considerations to concrete examples. We have two in hand. The first argument, predicting that I would like *The Wolf of Wall Street*, fits the pattern. Here's the argument again, for reference:

> *The Wolf of Wall Street* is directed by Martin Scorsese, and it stars Leonardo DiCaprio. Those two have collaborated several times in the past, such as on *Gangs of New York*, *The Aviator*, *The Departed*, and *Shutter Island*. I liked each of those movies, so I predict that I will like *The Wolf of Wall Street*.

The conclusion is something like "I will like *The Wolf of Wall Street*." Putting it that way, and looking at the general form of the conclusion of analogical arguments (*C* has Q), it's tempting to say that "*C"* designates me, while the property Q is something like "liking *The Wolf of Wall Street*." But that's not right. The thing that "*C*" designates has to be involved in the analogy in the first premise; it has to be *the thing* that's similar to the analogues. The analogy that this argument hinges on is between the various movies. It's not me that "*C*" corresponds to; it's the movie we're making the prediction about. *The Wolf of Wall Street* is what "*C*" picks out.

What property are we predicting it will have? Something like "liked by me." The analogues, a1 and so on in the schema, are the other movies: *Gangs of New York*, *The Aviator*, *The Departed*, and *Shutter Island*. These we know have the property Q (liked by me): I had already seen and liked these movies. That's the second premise: that the analogues have Q.

Finally, the first premise, which establishes the analogy among all the movies. What do they have in common? They were all directed by Martin Scorsese, and they all starred Leonardo DiCaprio. Those are the Ps—the properties they all share. P1 is "directed by Scorsese"; P2 is "stars DiCaprio."

The second argument we considered, about eating pork, also fits the pattern. Here it is again, for reference:

> Eating pork is immoral. Pigs are just as smart, cute, and playful as dogs and dolphins. Nobody would consider eating those animals. So why are pigs any different?

Again, looking at the conclusion—"Eating pork is immoral"—and looking at the general form of conclusions for analogical arguments—*C* has Q—it's tempting to just read off from the syntax of the sentence that "*C*" stands for "eating pork" and Q for "is immoral." But again, that's not right. Focus on the analogy: what *things* are being compared to one another? Answer: It's the animals: pigs, dogs, and dolphins; those are our *a*s and *C*. To determine which one is denoted by "*C*," we ask which animal is involved in the conclusion. Answer: It's pigs; they are denoted by "*C*." So we have to paraphrase our conclusion so that it fits the form "*C*" has Q. Something like "Pigs shouldn't be eaten" would work.

Q is the property "shouldn't be eaten," and the analogues are dogs and dolphins. They clearly have the property in question: as the argument notes, (most) everybody agrees they shouldn't be eaten. This is the second premise.

The first premise establishes the analogy. What do pigs have in common with dogs and dolphins? Answer: They're smart, cute, and playful. P1 = "is smart"; P2 = "is cute"; and P3 = "is playful."

The Evaluation of Analogical Arguments

Unlike in the case of deduction, we will not have to learn special techniques to use when evaluating these sorts of arguments. It's something we already know how to do, something we typically do automatically and unreflectively. The purpose of this section, then, is not to learn a new skill, but rather subject a familiar practice to critical scrutiny. We evaluate analogical arguments all the time without thinking about how we do it. We want to achieve a metacognitive perspective on the practice of evaluating arguments from analogy.

> **Metacognitive perspective:** Carefully thinking *about* a type of thinking that we typically engage in without much conscious deliberation.

We want to identify the criteria that we rely on to evaluate analogical reasoning—criteria that we apply without necessarily realizing that we're applying them. Achieving such metacognitive awareness is useful insofar as it makes us more self-aware, critical, and therefore effective reasoners.

Analogical arguments are inductive arguments. They give us reasons that are supposed to make their conclusions more probable. How probable, exactly? That's very hard to say. How probable was it that I would like *The Wolf of Wall Street* given that I had liked the other four Scorsese/DiCaprio collaborations? *I don't know*. How probable is it that it's wrong to eat pork, given that it's wrong to eat dogs and dolphins? *I really don't know*. It's hard to imagine how you would even begin to answer that question.

As we mentioned, while it's often impossible to evaluate inductive arguments by giving a *precise* probability of its conclusion, it is possible to make relative judgments about strength and weakness. Recall, new information can change the probability of the conclusion of an inductive argument. We can make *relative judgments* like this:

> If we add this new information as a premise, the new argument is stronger/weaker than the old argument; that is, the new information makes the conclusion more/less likely.

It is these types of relative judgments that we make when we evaluate analogical reasoning. We compare different arguments—with the difference being new information in the form of an added premise, or a different conclusion supported by the same premises—and judge one to be stronger or weaker than the other. Subjecting this practice to critical scrutiny, we can identify **six criteria** that we use to make such judgments.

We're going to be making relative judgments, so we need a baseline argument against which to compare others. Here is such an argument:

> Alice has taken four philosophy courses during her time in college. She got an A in all four. She has signed up to take another philosophy course this semester. I predict she will get an A in that course, too.

This is a simple argument from analogy, in which the future is predicted based on past experience. It fits the schema for analogical arguments: the new course she has signed up for is designated by "*C*"; the property we're predicting it has (Q) is that it is a course Alice will get an A in; the analogues are the four previous courses she's taken; what they have in common with the new course (P1) is that they are also philosophy classes; and they all have the property Q—Alice got an A in each.

Anyway, how strong is the baseline argument? How probable is its conclusion in light of its premises? I have no idea. It doesn't matter. We're now going to consider tweaks to the argument, and the effect that those will have on the probability of the conclusion. That is, we're going to consider slightly different arguments, with either new information added to the original premises or changes made to the prediction, and ask whether these altered new arguments are stronger or weaker than the baseline argument. This will reveal the six criteria that we use to make such judgments. We'll consider one criterion at a time.

1st Criteria: Number of Analogues

Suppose we alter the original argument by changing the number of prior philosophy courses Alice has taken. Instead of Alice having taken four philosophy courses before, we'll now suppose she has taken 14. We'll keep everything else about the argument the same: she got an A in all of them, and we're predicting she'll get an A in the new one. Are we more or less confident in the conclusion—the prediction of an A—with the altered premise? Is this new argument stronger or weaker than the baseline argument? Answer: It's stronger!

We've got Alice getting an A 14 times in a row instead of only four. That clearly makes the conclusion more probable. (How much more? Again, it doesn't matter.) What we did in this case is add more analogues. This reveals a general rule:

> Other things being equal, the more analogues in an analogical argument, the stronger the argument (and conversely, the fewer analogues, the weaker).

The number of analogues is one of the criteria we use to evaluate arguments from analogy.

2nd Criteria: Variety of Analogues

You'll notice that the original argument doesn't give us much information about the four courses Alice succeeded in previously and the new course she's about to take. All we know is that they're all philosophy courses. Suppose we tweak things. We're still in the dark about the new course Alice is about to take, but we know a bit more about the other four: one was a course in ancient Greek philosophy; one was a course on contemporary ethical theories; one was a course in formal logic; and the last one was a course on the philosophy of mind. Given this new information, are we more or less confident that she will succeed in the new course, whose topic is unknown to us? Is the argument stronger or weaker than the baseline argument? Answer: It is stronger. We don't know what kind of philosophy course Alice is about to take, but this new information gives us an indication that it doesn't really matter. She was able to succeed in a wide variety of courses, from mind to logic, from ancient Greek to contemporary ethics. This is evidence that Alice is good at philosophy generally, so that no matter what kind of course she's about to take, she'll probably do well in it. Again, this points to a general principle about how we evaluate analogical arguments:

> Other things being equal, the more variety there is among the analogues, the stronger the argument (and conversely, the less variety, the weaker).

3rd Criteria: Number of Similarities

In the baseline argument, the only thing the four previous courses and the new course have in common is that they're philosophy classes. Suppose we change that. Our newly tweaked argument predicts that Alice will get an A in the new course, which, like the four she succeeded in before, is cross-listed in the Department of Religious Studies and covers topics related to the philosophy of religion. Given this new information—that the new course and the four older courses were similar in ways we weren't aware of—are we more or less confident in the prediction that Alice will get another A? Is the argument stronger or weaker than the baseline argument? Answer: Again, it is stronger. Unlike the last example, this tweak gives us new information both about the four previous courses and the new one. The upshot of that information is that the courses are more similar than we knew; that is, they have more properties in common. To P1 = "is a philosophy course," we can add P2 = "is cross-listed with religious studies" and P3 = "covers topics in philosophy of religion." The more properties things have in common, the stronger the analogy between them. The stronger the analogy, the stronger the argument based on that analogy. We now know not just that Alice did well in not just in philosophy classes, but specifically in classes covering the philosophy of religion, and we know that the new class she's taking is also about the philosophy of religion. I'm much more confident predicting she'll do well again than I was when all I knew was that all the classes were philosophy; the new one could've been in a different topic that she wouldn't have liked. General principle:

> Other things being equal, the more properties involved in the analogy—the more similarities between the item in the conclusion and the analogues—the stronger the argument (and conversely, the fewer properties, the weaker).

4th Criteria: Number of Differences

An argument from analogy is built on the foundation of the similarities between the analogues and the target item in the conclusion—the analogy. Anything that weakens that foundation weakens the argument. So, to the extent that there are differences among those items, the argument is weaker.

Suppose we add new information to our baseline argument: the four philosophy courses Alice did well in before were all courses on the philosophy of mind; the new course is about the history of ancient Greek philosophy. Given this new information, are we more or less confident that she will succeed in the new course?

Is the argument stronger or weaker than the baseline argument? Answer: Clearly, the argument is weaker. The new course is on a completely different topic than the other ones. She did well in four straight philosophy of mind courses, but ancient Greek philosophy is quite different. I'm less confident that she'll get an A than I was before. If I add more differences, the argument gets even weaker. Supposing the four philosophy of mind courses were all taught by the same professor (the person in the department whose expertise is in that area), but the ancient Greek philosophy course is taught by someone different (the department's specialist in that topic). Different subject matter, different teachers: I'm even less optimistic about Alice's continued success. Generally speaking,

> Other things being equal, the more differences there are between the analogues and the item in the conclusion, the weaker the argument from analogy.

5th Criteria: Relevance of Similarities and Differences

Not all similarities and differences are capable of strengthening or weakening an argument from analogy, however. Suppose we tweak the original argument by adding the new information that the new course and the four previous courses all have their weekly meetings in the same campus building. This is an additional property that the courses have in common, which, as we just saw, other things being equal, should strengthen the argument. But other things are not equal in this case. That's because it's very hard to imagine how the location of the classroom would have anything to do with the prediction we're making—that Alice will get an A in the course. Classroom location is simply not relevant to success in a course. Therefore, this new information does not strengthen the argument. Nor does it weaken it; I'm not inclined to doubt that Alice will do well considering the information about location. It simply has no effect at all on my appraisal of her chances. Similarly, if we tweak the original argument to add a difference between the new class and the other four—for example, that all four previous classes were in the same building, while the new one is in a different building—there is no effect on our confidence in the conclusion. Again, the building in which a class meets is generally not relevant to how well someone does.

Contrast these cases with the new information that the new course and the previous four are all taught by the same professor. Now that strengthens the argument! Alice has gotten an A four times in a row from this professor—all the more reason to expect she'll receive another one. This tidbit strengthens the argument because the new similarity—the same person teaches all the courses—is relevant to the prediction we're making—that Alice will do well. Who teaches a class can make a difference to how students do—either because they're easy graders, or because they're great teachers, or because the student and the teacher are in tune with one another, etc. Even a difference between the analogues and the item in the conclusion, with the right kind of relevance, can strengthen an argument. Suppose the other four philosophy classes were taught by the same teacher, but the new one is taught by a TA—who just happens to be her boyfriend. That's a difference, but one that makes the conclusion—that Alice will do well—more probable. However, in such a case, we are more apt to treat this as an additional "stand-alone" premise, one without real influence on the strength or weakness of the analogy. Generally speaking, Careful attention must be paid to the relevance of any similarities and differences to the property in the conclusion; the effect on strength varies.

6th Criteria: Modesty/Ambition of the Conclusion

Suppose we leave everything about the premises in the original baseline argument the same: four philosophy classes, an A in each, new philosophy class. Instead of adding to that part of the argument, we'll tweak the conclusion:

> Instead of predicting that Alice will get an A in the class, we'll predict that she'll simply *pass* the course.

Are we more or less confident that this prediction will come true? Is the new, tweaked argument stronger or weaker than the baseline argument? Answer: It's stronger. We are more confident in the prediction that Alice will pass than we are in the prediction that she will get another A, for the simple reason that it's much easier to pass than it is to get an A. That is, the prediction of passing is a much more modest prediction than the prediction of an A.

Suppose we tweak the conclusion in the opposite direction—not more modest, but more ambitious. Alice has gotten an A in four straight philosophy classes. She's about to take another one, and I predict that she will do so well that her professor will suggest that she publish her term paper in one of the most prestigious philosophical journals, and that she will be offered a three-year research fellowship at the Institute for Advanced Study at Princeton University. That's a bold prediction! Meaning, of course, that it's very unlikely to happen. Getting an A is one thing; getting an invitation to be a visiting scholar at one of the most prestigious academic institutions in the world is quite another. The argument with this ambitious conclusion is weaker than the baseline argument. General principle:

> The more modest the argument's conclusion, the stronger the argument; the more ambitious, the weaker.

Table of Six Criteria for Evaluating Arguments by Analogy

Criteria #1: **Number of Analogues**—Other things being equal, the more analogues in an analogical argument, the stronger the argument (and conversely, the fewer analogues, the weaker).

Criteria #2: **Variety of Analogues**—Other things being equal, the more variety there is among the analogues, the stronger the argument (and conversely, the less variety, the weaker).

Criteria #3: **Number of Similarities**—Other things being equal, the more properties involved in the analogy—the more similarities between the item in the conclusion and the analogues—the stronger the argument (and conversely, the fewer properties, the weaker).

Criteria #4: **Number of Differences**—Other things being equal, the more differences there are between the analogues and the item in the conclusion, the weaker the argument from analogy.

Criteria #5: **Relevance of Similarities and Differences**—Careful attention must be paid to the relevance of any similarities and differences to the property in the conclusion; the effect on strength varies.

Criteria #6: **Modesty/Ambition of the Conclusion**—The more modest the argument's conclusion, the stronger the argument; the more ambitious, the weaker.

Refutation by Analogy

We can use arguments from analogy for a specific logical task: refuting someone else's argument, showing that we are not compelled to accept the conclusion. Recall the case of deductive arguments. To refute those meant showing they are invalid—we had to produce a counterexample. This meant providing a new argument with the same logical form as the original, the same pattern of syntax found in the premises and conclusion. However, this new argument needed to be obviously invalid, in that its premises were in fact true and its conclusion in fact false.

We can use a similar procedure to refute inductive arguments. Of course, the standard of evaluation is different for induction: we don't judge them according to the black and white standard of validity. And as a result, our judgments have *less to do with form than with content*. Nevertheless, refutation along similar lines is possible, and analogies are the key to the technique.

To refute an inductive argument, we produce a new argument that's obviously bad—just as we did in the case of deduction. We don't have a precise notion of logical form for inductive arguments, so we can't demand that the refuting argument have the same syntax as the original; rather, we want the new argument to have *an analogous form of reasoning* to the original. The stronger the analogy between the refuting and refuted arguments, the more decisive the refutation. We cannot produce the kind of knock-down refutations that were possible in the case of deductive arguments, where the standard of evaluation—validity—does not admit of degrees of goodness or badness, but the technique can be quite effective.

Consider the following scenario:

You and your friend are watching TV. A news story reports that *Duck Dynasty* star Willie Robertson said he supports Trump because both of them have been successful businessmen and stars of reality TV shows. Your friend scoffs and says:

> By that logic, does that mean Hugh Hefner's success with *Playboy* and his occasional appearances on *Bad Girls Club* warrant him as a worthy president?

Your friend is refuting the argument of Willie Robertson, the *Duck Dynasty* star. Robertson's argument is something like this:

> Trump is a successful businessman and reality TV star.
>
> Therefore, he would be a good president.

To refute this, your friend makes an analogy between Trump and Hugh Hefner. Your friend uses this to produce an analogous argument:

> Hugh Hefner is a successful businessman and reality TV star.
>
> Therefore, Hugh Hefner would make a good president
>
> This is regarded by your friend as obviously bad.
>
> Therefore, the original argument about Trump is just as obviously bad.

What makes it obviously bad is that it has a conclusion that nobody would agree with; nobody is likely to believe that Hugh Hefner would make a good president. That's how these refutations work. They attempt to demonstrate that the original argument is lousy by showing that *you can use the same or very similar reasoning* to arrive at an absurd conclusion. That is, they use an analogy between the two attempts at reasoning to demonstrate the flaw in the original argument (which is the target we want to pin "bad reasoning" onto).

Here's another example, from a group called "Iowans for Public Education," who wanted to support funding for public schools by showing flaws in the views of people who want to use those funds for private schooling options.

> Next to a picture of a seemingly wealthy lady is the following text:
>
> "My husband and I have decided the local parks just aren't good enough for our kids. We'd rather use the country club, and we are hoping state tax dollars will pay for it. We are advocating for Park Savings Accounts, or PSAs. We promise to no longer use the local parks. To hell with anyone else or the community as a whole. We want our tax dollars to be used to make the best choice for our family."
>
> Sound ridiculous? Tell your legislator to vote NO on Education Savings Accounts (ESAs), a.k.a. school vouchers.

The argument that Iowans for Public Education put in the mouth of the lady on the poster is meant to refute reasoning used by advocates for "school choice." Those advocates say that they ought to have the right to opt out of public education and keep the tax dollars they would otherwise pay for public schools to pay for their kids to attend private schools. A similar line of reasoning sounds pretty crazy when you replace "public schools" with "public parks," "private schools" with "country clubs," and "ESAs" with "PSAs." The argument from analogy works here because our judgment of the line of reasoning in the parks argument is now attached to the target: the reasoning offered in the school-choice arguments.

Since these sorts of refutations rely on analogies, they are only as strong as the analogy between the refuting and refuted arguments. There is room for dispute on that question. Advocates for school vouchers might point out that schools and parks are completely different things, that schools are much more important to the future prospects of children, and that given the importance of education, families should have the right to choose what they think is best. Or something like that. The point is, when attacked by an argument from analogy, those defending their position will need to undermine the strength of the analogy. Their strategy will be to chip away at the alleged similarities between analogue and target. Those pushing for the analogy will try to counter-respond by maintaining enough of the criteria for evaluating arguments from analogy (e.g., Number Analogues, Relevance of Similarities, etc.). This is always a complex affair. The kinds of knock-down refutations that were possible for deductive arguments are not possible for inductive arguments.

Exercises

1. Show how the following arguments fit the abstract schema for arguments from analogy:

   a1, a2, …, and *C* all have P1, P2, …
   a1, a2, …, all have Q ____________________
   So *C* has Q

   a. You should really eat at Papa Giorgio's; you'll love it. It's just like Mama DiSilvio's and Matteo's, which I know you love: they serve old-fashioned Italian-American food, they have a laid-back atmosphere, and the wine list is extensive.
   b. George R.R. Martin deserves to rank among the greats in the fantasy literature genre. Like C.S. Lewis and J.R.R. Tolkien before him, he has created a richly detailed world, populated it with compelling characters, and told a tale that is not only exciting, but which features universal and timeless themes concerning human nature.
   c. Yes, African Americans are incarcerated at higher rates than whites. But blaming this on systemic racial bias in the criminal justice system is absurd. That's like saying the NBA is racist because there are more black players than white players, or claiming that the medical establishment is racist because African Americans die young more often.

2. Consider the following baseline argument:

   I've taken vacations to Florida six times before, and I've enjoyed each visit. I'm planning to go to Florida again this year, and I fully expect yet another enjoyable vacation.
   For each of the following changes:

   i. Decide which produces an argument that's weaker or stronger than the baseline argument and

ii. Indicate which of the six criteria for evaluating analogical arguments justifies that judgment

a. All of my trips were visits to Disney World, and this one will be no different.
b. In fact, I've vacationed in Florida 60 times and enjoyed every visit.
c. I expect that I will enjoy this trip so much I will decide to move to Florida.
d. On my previous visits to Florida, I've gone to the beaches; the theme parks; Everglades National Park; and various cities, from Jacksonville to Key West.
e. I've always flown to Florida on Delta Airlines in the past; this time I'm going on a United flight.
f. All of my past visits were during the winter months; this time I'm going in the summer.
g. I predict that I will find this trip more enjoyable than a visit to the dentist.
h. I've only been to Florida once before.
i. On my previous visits, I drove to Florida in my Dodge minivan, and I'm planning on driving the van down again this time.
j. All my visits have been to Daytona Beach for the Daytona 500; same thing this time.
k. I've stayed in beachside bungalows, big fancy hotels, time-share condominiums—even a shack out in the swamp.

3. For each of the following passages, explicate the argument being refuted and the argument or arguments doing the refuting.

a. Republicans tell us that, because at some point 40 years from now a shortfall in revenue for Social Security is projected, we should cut benefits now. Cut them now because we might have to cut them in the future? I've got a medium-sized tree in my yard. 40 years from now, it may grow so large that its branches hang over my roof. Should I chop it down?
b. Opponents of gay marriage tell us that it flies in the face of a tradition going back millennia: that marriage is between a man and a woman. There were lots of traditions that lasted a long time: the tradition that it was OK for some people to own other people as slaves, the tradition that women couldn't participate in the electoral process—the list goes on. That it's traditional doesn't make it right.
c. Some people claim that their children should be exempted from getting vaccinated for common diseases because the practice conflicts with their religious beliefs. But religion can't be used to justify just anything. If a Satanist tried to defend himself against charges of abusing children by claiming that such practices were a form of religious expression, would we let him get away with it?

Causal Reasoning

Inductive arguments are frequently used to support claims about cause and effect. These arguments come in a number of different forms. The most straightforward is what is called enumerative induction. This is an argument that makes a (non-hasty) generalization, inferring that one event or type of event causes another on the basis of a (large) number of particular observations of the cause immediately preceding the effect. To use a very famous example (from the history of philosophy, per David Hume, the 18th century Scottish philosopher), we can infer from observations of a number of billiard ball collisions that the first ball colliding with the second causes the second ball to move.

This is all well and good, so far as it goes. It just doesn't go very far. If we want to establish a robust knowledge of what causes the natural phenomena we're interested in, we need techniques that are more sophisticated than simple enumerative induction. Fortunately, there are such techniques. These are patterns of reasoning identified and catalogued by the 19th century English philosopher John Stuart Mill. The inferential forms Mill enumerated have come to be called "Mill's Methods," because he thought of them as tools to be used in the investigation of nature—methods of discovering the causes of natural phenomena. In this section, we will look at Mill's Methods each in turn (there are five of them), using examples to illustrate each. We will finish with a discussion of the limitations of the methods and the difficulty of isolating causes.

The Meaning(s) of "Cause"

Before we proceed, however, we must issue something of a disclaimer: when we say that one action or event causes another, we don't really know what the hell we're talking about. OK, maybe that's putting it a bit too strongly. The point is this: the meaning of "cause" has been the subject of intense philosophical debate since ancient times (in both Greece and India)—debate that continues to this day. Myriad philosophical theories have been put forth over the millennia about the nature of causation, and there is no general agreement about just what it is (or whether causes are even real!).

We're not going to wade into those philosophical waters; they're too deep. Instead, we'll merely dip our toes in, by making a preliminary observation about the word "cause"—an observation that gives some hint as to why it's been the subject of so much philosophical deliberation for so long. The observation is this:

> There are a number of distinct, but perfectly acceptable, ways that we use the word "cause" in everyday language. We attach *different incompatible meanings* to the term in different contexts.

Consider this scenario: I'm in my backyard vegetable garden with my younger daughter (age four at the time). She's "helping" me in my labors by watering some of the plants. She asks, "Daddy, why do we have to water the plants?" I might reply, "We do that because water causes the plants to grow." This is a perfectly ordinary claim about cause and effect; it is uncontroversial and true. What do I mean by "causes" in this sentence? I mean that water is **a necessary condition** for the plants to grow. Without water, there will be no growth. It is not a sufficient condition for plant growth, though: you also need sunlight, good soil, etc.

Consider another completely ordinary, uncontroversial truth about causation: decapitation causes death. What do I mean by "causes" in this sentence? I mean that decapitation is **a sufficient condition** for death. If death is the result you're after, decapitation will do the trick on its own; nothing else is needed. It is not (thank goodness) a necessary condition for death, however. There are lots of other ways to die besides beheading.

Finally, consider this true claim: smoking causes cancer. What do I mean by "causes" in this sentence? Well, *I don't mean* that smoking is a sufficient condition for cancer. Lots of people smoke all their lives but are lucky enough not to get cancer. Moreover, *I don't mean* that smoking is a necessary condition for cancer. Lots of people get cancer—even lung cancer—despite having never smoked. Rather, what I mean is that smoking **tends to produce** cancer; it increases the probability that one will get cancer.

So, we have three totally ordinary uses of the word "cause," with three completely different meanings: cause as **necessary** condition, **sufficient** condition, and mere **tendency** (neither necessary nor sufficient). These are incompatible, but all acceptable in their contexts. We could go on to list even more uses for the term, but the point has been made. Causation is a slippery concept, which is why philosophers have been struggling for so long to capture its precise meaning. In what follows, we will set aside these concerns and speak about cause and effect without hedging or disclaimers, but it's useful to keep in mind that doing so papers over some deep and difficult philosophical problems.

Mill's Methods

John Stuart Mill identified five different *patterns of reasoning* that one could use to discover causes. These are *argument forms*, the conclusions of which involve a claim to the effect that one thing causes (or is causally related to) another. They can be used alone or in combination, depending on the circumstances.

As was the case with analogical reasoning, these are patterns of inference that we already employ unreflectively in everyday life. The benefit of making them explicit and subjecting them to critical scrutiny is that we thereby achieve a **metacognitive perspective**—a perspective from which we can become more self-aware, effective reasoners. This is especially important in the context of causal reasoning, since, as we shall see, there are many pitfalls in this domain that we are prone to fall into, many common errors that people make when thinking about cause and effect.

Method of Agreement

We could sum up this reasoning pattern abstractly thus: We want to find the cause of a phenomenon: call it X. We examine a variety of circumstances in which X occurs, looking for potential causes. The circumstances differ in various ways, but they each have in common that they feature the same potential cause: call it A. We conclude that A causes X. Each of the circumstances agrees with the others in the sense that they all feature the same potential cause—hence, the Method of Agreement. Consider the following story:

I've been suffering from heartburn recently. Seems like at least two or three days a week, by about dinnertime, I've got that horrible feeling of indigestion in my chest and that yucky taste in my mouth. Acid reflux: ugh. I've got to do something about this. What could be causing my heartburn, I wonder? I know that the things you eat and drink are typical causes of the condition, so I start thinking back, looking at what I've consumed on the days when I felt bad. As I recall, all of the recent days on which I suffered heartburn were different in various ways: my dinners ranged from falafel to spaghetti to spicy burritos; sometimes I had a big lunch, sometimes very little; on some days I drank a lot of coffee at breakfast, but other days not any at all. But now that I think about it, *one thing stands out*: I've been in a nostalgic mood lately, thinking about the good old days when I was a carefree college student. I've been listening to lots of music from that time, watching old movies, etc. And as part of that trip down memory lane, I've re-acquired a taste for one of my favorite beverages from that era—Mountain Dew. I've been treating myself to a nice bottle of the stuff with lunch now and again. And sure enough, each of the days that I got heartburn was a day when I drank Mountain Dew at lunch. Huh. I guess the Mountain Dew is causing my heartburn. I better stop drinking it.

This little story is an instance of Mill's Method of Agreement. It's a pattern of reasoning that one can use to figure out the cause of some phenomenon of interest. In this case, the phenomenon I want to discover the cause of is my recent episodes of heartburn. I eventually figure out that the cause is Mountain Dew. Though the circumstances of my heartburn varied in many ways, consuming Mountain Dew was the common feature in each instance of heartburn. So, I conclude that it causes my heartburn.

In the story above, the phenomenon X that I wanted to find the cause of was heartburn; the various circumstances were the days on which I had suffered that condition, and they varied with respect to potential causes (foods and beverages consumed); however, they all agreed in featuring Mountain Dew, which is the factor A causing the heartburn, X.

More simply, we can sum up the Method of Agreement as a simple question:

Method of Agreement: What causal factor is present whenever the phenomenon of interest is present?

In the case of our little story, Mountain Dew was present whenever heartburn was present, so we concluded that it was the cause.

Method of Difference

We can sum up this pattern of reasoning abstractly thus: We want to find the cause of a phenomenon: call it X. We examine a variety of circumstances in which X occurs, looking for potential causes. The circumstances differ in various ways, but they each have in common that when we remove from them a potential cause—call it A—the phenomenon disappears. We conclude that A causes X. If we introduce the same difference in all of the circumstances—removing the causal factor—we see the same effect—disappearance of the phenomenon. Hence, the Method of Difference. Consider the following story:

> Everybody in my house has a rash! Itchy skin, little red bumps; it's annoying. It's not just the grownups—me and my wife—but the kids, too. Even the dog has been scratching herself constantly! What could possibly be causing our discomfort? My wife and I brainstorm, and she remembers that she recently changed brands of laundry detergent. Maybe that's it. So we re-wash all the laundry (including the pillow that the dog sleeps on in the windowsill) in the old detergent and wait. Sure enough, within a day or two, everybody's rash is gone. Sweet relief!

This story presents an instance of Mill's Method of Difference. Again, we use this pattern of reasoning to discover the cause of some phenomenon that interests us—in this case, the rash we all have. We end up discovering that the cause is the new laundry detergent. We isolated this cause by removing that factor and seeing what happened.

In our story, the phenomenon we wanted to explain, X, was the rash. The varying circumstances are the different inhabitants of my house—Mom, Dad, kids, even the dog—and the different factors affecting them. The factor that we removed from each, A, was the new laundry detergent. When we did that, the rash went away, so the detergent was the cause of the rash—A caused X.

More simply, we can sum up the Method of Difference as a simple question:

Method of Difference: What causal factor is absent whenever the phenomenon of interest is absent?

In the case of our little story, when the detergent was absent, so too was the rash. We concluded that the detergent caused the rash.

Joint Method of Agreement and Difference

This isn't really a new method at all. It's just a combination of the first two. The Methods of Agreement and Difference are complementary; each can serve as a check on the other. Using them in combination is an extremely effective way to isolate causes.

The Joint Method is an important tool in many forms of research. It's the pattern of reasoning used in what we call controlled studies. In such a study, we split our subjects into two groups to verify casual claims. An example shows how this works. Suppose I've formulated a pill that I think is a miracle cure for baldness. *I'm* *gonna be rich!* But first, I need to see if it really works. So I gather a bunch of bald men together for a controlled study. One group gets the actual drug; the other, the control group, gets a sugar pill—not the real drug at all, but a mere placebo, a "fake" pill that we know won't have any causal influence on the test subjects. Then I wait and see what happens. If my drug is a good as I think it is, two things will happen: first, the group that got the drug will grow new hair; second, the group that got the placebo won't grow new hair. If either of these things fails to happen, it's back to the drawing board. Obviously, if the group that got the drug didn't get any new hair, my baldness cure is a dud. But in addition, if the group that got the mere placebo grew new hair, then *something else* besides my drug has to be the cause. In this way we rule out a false positive result that we would have gotten if I only had one group of men take my new pill and they saw their hair growth improve—it would falsely look like my pill was the cause of that new hair growth.

Both the Method of Agreement and the Method of Difference are being used in a controlled study. I'm using the Method of Agreement on the group that got the drug. I'm hoping that whenever the causal factor (my miracle pill) is present, so too will be the phenomenon of interest (hair growth). The control group complements this with the Method of Difference. For them, I'm hoping that whenever the causal factor (the miracle pill) is absent, so too will be the phenomenon of interest (hair growth). If both things happen, I've got strong confirmation that my drug causes hair growth.

Joint Method of Agreement and Disagreement: Use of both Agreement and Disagreement to check results against one another.

Method of Residues

"Residue" in this context just means the remainder, that which is left over. This pattern of reasoning, put abstractly, runs something like this: We observe a series of phenomena, call them X1, X2, X3, ..., X*n*. As a matter of background knowledge, we know that X1 is caused by A1, that X2 is caused by A2, and so on. But when we exhaust our background knowledge of the causes of phenomena, we're left with one, X*n*, that is inexplicable in those terms. So, we must seek out an additional causal factor, A*n*, as the cause of X*n*. The leftover phenomenon, X*n*, inexplicable in terms of our background knowledge, is the residue. Consider the following story. You may not be familiar with all the terminology, but just pay close attention to the numbers and see how they add up...

> I'm running a business. Let's call it LogiCorp. For a modest fee, the highly trained logicians at LogiCorp will evaluate all of your deductive arguments, issuing Certificates of Validity (or Invalidity) that are legally binding in all fifty states. Satisfaction guaranteed. Anyway, as should be obvious from

that brief description of the business model, LogiCorp is a highly profitable enterprise. But last year's results were disappointing. Profits were down 20% from the year before.

Some of this was expected. We undertook a renovation of the LogiCorp World Headquarters that year, and the cost had an effect on our bottom line: half of the lost profits, 10%, can be chalked up to the renovation expenses. Also, as healthcare costs continue to rise, we had to spend additional money on our employees' benefits packages; these expenditures account for an additional 3% of profit shortfall. Finally, another portion of the drop in profits can be explained by the entry of a competitor into the marketplace. The upstart firm Arguments R Us, with its fast turnaround times and ultra-cheap prices, has been cutting into our market share. Their services are totally inferior to ours (you should see the shoddy shading technique in their Venn diagrams!) and LogiCorp will crush them eventually, but for now they're hurting our business: competition from Arguments R Us accounts for a 5% drop in our profits.

As CEO, I was of course aware of all these potential problems throughout the year, so when I looked at the numbers at the end, I wasn't surprised. But when I added up the contributions from the three factors I knew about—10% from the renovation, 3% from the healthcare expenditures, 5% from outside competition—I came up short. Those causes only account for an 18% shortfall in profit, but we were down 20% on the year; there was an extra 2% shortfall that I couldn't explain.

I'm a suspicious guy, so I hired an outside security firm to monitor the activities of various highly placed employees at my firm. And I'm glad I did! Turns out my Chief Financial Officer had been taking lavish weekend vacations to Las Vegas and charging his expenses to the company credit card. His thievery surely accounts for the extra 2%. I immediately fired the jerk.

This little story presents an instance of Mill's Method of Residues. In our story, that was the additional 2% profit shortfall that couldn't be explained in terms of the causal factors we were already aware of, namely the headquarters renovation (A1, which caused X1, a 10% shortfall), the healthcare expenses (A2, which caused X2, a 3% shortfall), and the competition from Arguments R Us (A3, which caused X3, a 5% shortfall). We had to search for another, previously unknown cause for the final, residual 2%.

Method of Residues: Account for as many phenomena as possible with known causes until only residual phenomena remain with unaccounted causes. Focus on candidates to explain the residual phenomena.

Method of Concomitant Variation

Put abstractly, this pattern of reasoning goes something like this: We observe that, holding other factors constant, an increase or decrease in some causal factor A is always accompanied by a corresponding increase or decrease in some phenomenon X. We conclude that A and X are causally related.

Things that "vary concomitantly" are things, to put it more simply, that change together. As A changes—goes up or down—X changes, too. There are two ways things can vary concomitantly: directly or inversely. If A and X vary directly, that means that an increase in one will be accompanied by an increase in the other (and a decrease in one will be accompanied by a decrease in the other); if A and X vary inversely, that means an increase in one will be accompanied by a decrease in the other.

Fact: if you're a person who currently maintains a fairly steady weight, and you change nothing else about your lifestyle, adding 500 calories per day to your diet will cause you to gain weight. Conversely, if you cut 500 calories per day from your diet, you will lose weight. That is, calorie consumption and weight are causally related: consuming more will cause weight gain; consuming less will cause weight loss.

Another fact: if you're a person who currently maintains a steady weight, and you change nothing else about your lifestyle, adding an hour of vigorous exercise per day to your routine will cause you to lose weight. Conversely (assuming you already exercise a heck of a lot), cutting that amount of exercise from your routine will cause you to gain weight. That is, exercise and weight are causally related: exercising more will cause weight loss; exercising less will cause weight gain.

I know about the cause-and-effect relationships above because of the Method of Concomitant Variation. In our first example, calorie consumption (A) and weight (X) *vary directly*. As calorie consumption increases, weight increases; as calorie consumption decreases, weight decreases. In our second example, exercise (A) and weight (X) *vary inversely*. As exercise increases, weight decreases; as exercise decreases, weight increases. Either way, when things change together in this way, when they vary concomitantly, we conclude that they are causally related.

Method of Concomitant Variation: Attribute a specific cause to some phenomenon when we find *its* increase or decrease is always accompanied by a *corresponding* increase or decrease in the *phenomena*.

Mill's Methods for Discovering Causes

Method of Agreement: Pursue the Q: What causal factor is present whenever the phenomenon of interest is present?

Method of Difference: Pursue the Q: What causal factor is absent whenever the phenomenon of interest is absent?

Joint Method of Agreement and Difference: Use of both Agreement and Disagreement to check results against one another.

Method of Residues: Account for as many phenomena as possible with known causes until only residual phenomena remain with unaccounted causes. Focus on candidates to explain the residual phenomena.

Method of Concomitant Variation: Attribute a specific cause to some phenomenon when we find *its* increase or decrease is always accompanied by a *corresponding* increase or decrease in the *phenomena*.

The Difficulty of Isolating Causes

Mill's Methods are useful in discovering the causes of phenomena in the world, but their usefulness should not be overstated. Unless they are employed thoughtfully, they can lead an investigator astray.

A classic example of this is the parable of the drunken logician.[1] After a long day at work on a Monday, a certain logician heads home wanting to unwind. So he mixes himself a "7 and 7"—Seagram's 7 Crown whiskey and 7-Up. It tastes so good, he makes another—and another, and another. He drinks seven of these cocktails, passes out in his clothes, and wakes up feeling terrible (headache, nausea, etc.). On Tuesday, after dragging himself into work, toughing it through the day, then finally getting home, he decides to take the edge off with a different drink: brandy and 7-Up. He gets carried away again, and ends up drinking seven of these cocktails, with the same result: passing out in his clothes and waking up feeling awful on Wednesday. So, on Wednesday night, our logician decides to mix things up again: scotch and 7-Up. He drinks seven of these: same results. But he perseveres! Thursday night, it's seven vodka and 7-Ups; another blistering hangover on Friday. So on Friday at work, he sits down to figure out what's going on.

He's got a phenomenon—hangover symptoms every morning of that week—that he wants to discover the cause of. He's a professional logician, intimately familiar with Mill's Methods, so he figures he ought to be able

1. See Copi and Cohen, p. 547

to discover the cause. He looks back at the week and uses the Method of Agreement, asking, "What factor was present every time the phenomenon was?" He concludes that the cause of his hangovers is 7-Up.

Our drunken logician applied the Method of Agreement correctly: 7-Up *was* indeed present every time. But it clearly wasn't the cause of his hangovers. The lesson is that Mill's Methods are useful tools for discovering causes, but their results are not always definitive. Uncritical application of the methods can lead one astray. Additional knowledge must often be brought to bear on our conclusions (e.g., knowledge of the alcohol content of 7-Up would have probably helped our logician avoid his error). Critical evaluation of a hypothesis requires more than the sole application of Mill's Methods.

This is especially true of the Method of Concomitant Variation. You may have heard the old saw that "correlation does not imply causation." It's useful to keep this corrective in mind when using the Method of Concomitant Variation. That two things vary concomitantly is a hint that they may be causally related, but it is not definitive proof that they are. They may be *separate effects* of a different, unknown cause; they may be completely causally unrelated.

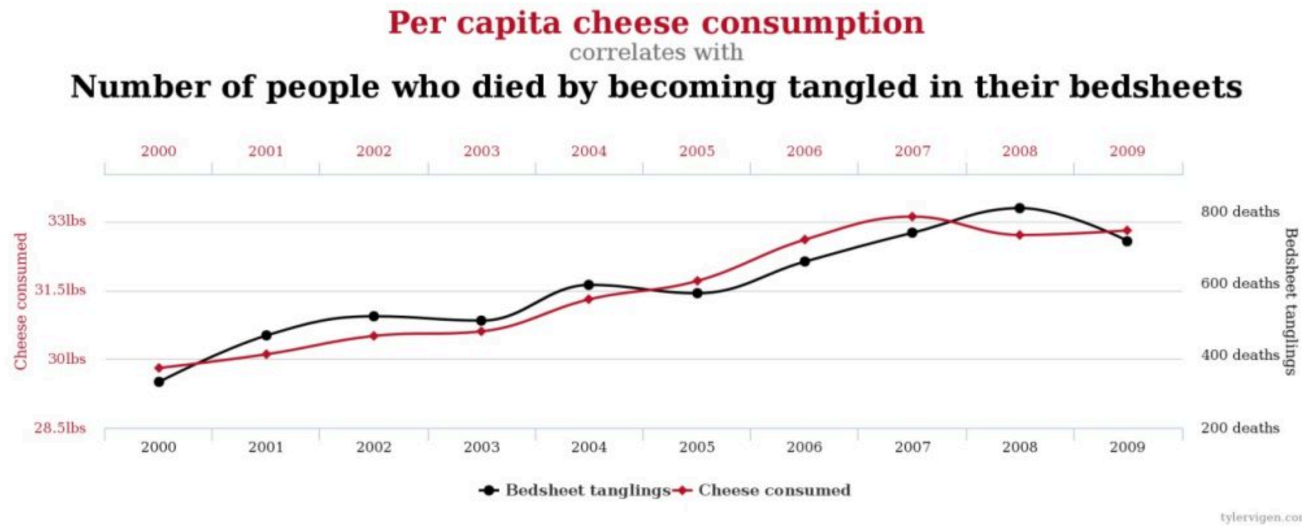
It is true, for example, that among children, shoe size and reading ability vary directly: children with bigger feet are better readers than those with smaller feet. Wow! So large feet cause better reading? Of course not. Larger feet and better reading ability are both effects of the same cause: *getting older*. Older kids wear bigger shoes than younger kids, and they also do better on reading tests. Duh.

It is also true, for example, that hospital quality and death rate vary directly: that is, the higher quality the hospital (prestige of doctors, training of staff, sophistication of equipment, etc.), on average, the higher the death rate at that hospital. That's counterintuitive! Does that mean that high hospital quality causes high death rates? Of course not. Better hospitals have higher mortality rates because the extremely sick, most badly injured patients are taken to those hospitals, rather than to the ones with lower-quality staff and equipment. Alas, these people die more often, but not because they're at a good hospital; it's exactly the reverse.

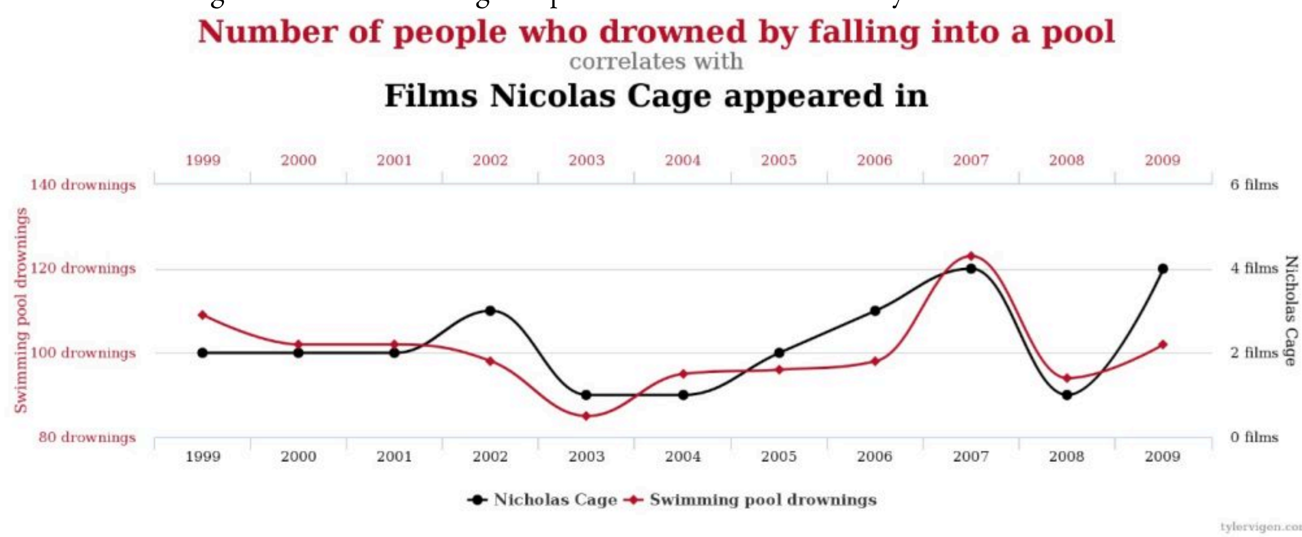
Spurious correlations—those that don't involve any causal connection at all—are easy to find in the age of "big data." With publicly available databases archiving large amounts of data, and computers with the processing power to search them and look for correlations, it is possible to find many examples of phenomena that vary concomitantly but are obviously not causally connected.

A very clever person named Tyler Vigen set about doing this and created a website where he posted his (often very amusing) discoveries.[2] For example, he found that between 2000 and 2009, per capita cheese consumption among Americans was very closely correlated with the number of deaths caused by people becoming entangled in their bedsheets.

2. http://tylervigen.com/spurious-correlations The site has a tool that allows the user to search for correlations.



These two phenomena vary directly, but it's hard to imagine how they could be causally related. It's even more difficult to imagine how the following two phenomena could be causally related:



So, Mill's Methods can't just be applied willy-nilly; one could end up "discovering" causal connections where none exist. They can provide clues as to potential causal relationships, but care and critical analysis are required to confirm those results. It's important to keep in mind that the various methods can work in concert, providing a check on each other. If the drunken logician, for example, had applied the Method of Difference—removing the 7-Up but keeping everything else the same—he would have discovered his error (he would've kept getting hangovers). The combination of the Methods of Agreement and Difference—the Joint Method, the controlled study—is an invaluable tool in modern scientific research. A properly conducted controlled study can provide quite convincing evidence of causal connections (or a lack thereof).

Of course, properly conducting a controlled study is not as easy as it sounds. It involves more than just the application of the Joint Method of Agreement and Difference. There are other potentially confounding factors that must be accounted for in order for such a study to yield reliable results. For example, it's important to take great care in separating subjects into the test and control groups: there can be no systematic difference between the two groups other than the factor that we're testing; if there is, we cannot say whether the factor we're testing or the difference between the groups is the cause of any effects observed.

Suppose we were conducting a study to test the common belief that vitamin C is effective in treating the common cold. We gather 100 subjects experiencing the onset of cold symptoms. We want one group of 50 to get vitamin C supplements, and one group of 50—the control group—not to receive them.

How do we decide who gets placed into which group? We could ask for volunteers. But doing so might create a systematic difference between the two groups. People who hear "vitamin C" and think, "*yeah, that's the group for me*" might be people who are more inclined to eat fruits and vegetables, for example, and might therefore be healthier on average than people who are turned off by the idea of receiving vitamin C supplements. This difference between the groups might lead to different results in how their colds progress.

Instead of asking for volunteers, we might just assign the first 50 people who show up to the vitamin C group, and the last 50 to the control group. But this could lead to differences, as well. The people who show up earlier might be early risers, who might be healthier on average than those who straggle in late.

The best way to avoid systematic differences between test and control groups is to randomly assign subjects to each. We refer to studies conducted this way as randomized controlled studies. And besides randomization, other measures can be taken to improve reliability. The best kinds of controlled studies are "double-blind." This means that neither the subjects nor the people conducting the study know which group is the control and which group is receiving the actual treatment. (This information is hidden from the researchers only while the study is ongoing; they are told later, of course, so they can interpret the results.) This measure is necessary because of the psychological tendency for people's observations to be biased based on their expectations.

For example, if the control group in our vitamin C experiment knew they were not getting any treatment for their colds, they might be more inclined to report that they weren't feeling any better. Conversely, if the members of the group receiving the vitamin supplements knew that they were getting treated, they might be more inclined to report that their symptoms weren't as bad. This is why the usual practice is to keep subjects in the dark about which group they're in, giving a placebo to the members of the control group. It's important to keep the people conducting the study "blind" for the same reasons. If they knew which group was which, they might be more inclined to observe improvement in the test group and a lack of improvement in the control group. In addition, in their interactions with the subjects, they may unknowingly give away information about which group was which via subconscious signals.

Hence, the gold standard for medical research (and other fields) is the double-blind controlled study. It's not always possible to create those conditions—sometimes the best doctors can do is to use the Method of Agreement and merely note commonalities amongst a group of patients suffering from the same condition, for example—but the most reliable results come from such tests. Discovering causes is hard in many contexts. Mill's Methods are a useful starting point, and they accurately model the underlying inference patterns involved in such research, but in practice they must be supplemented with additional measures and analytical rigor in order to yield definitive results. They can give us clues about causes, but they aren't definitive evidence. Remember, these are inductive, not deductive arguments.

Exercises

1. What is meant by the word "cause" in the following— indicate if "cause" means necessary condition, sufficient condition, or mere tendency?
    a. Throwing a brick through a window causes it to break.
    b. Slavery caused the American Civil War.
    c. Exposure to the cold causes frostbite.
    d. Running causes knee injuries.
    e. Closing your eyes causes you not to be able to see.

2. Consider the following scenario and answer the questions about it:

    Alfonse, Bertram, Claire, Dominic, Ernesto, and Francine all go out to dinner at a local greasy spoon. There are six items on the menu: shrimp cocktail, mushroom/barley soup, burger, fries, steamed carrots, and ice cream. This is what they ate:

    - Alfonse: shrimp, soup, fries
    - Bertram: burger, fries, carrots, ice cream
    - Claire: soup, burger, fries, carrots
    - Dominic: shrimp, soup, fries, ice cream
    - Ernesto: burger, fries, carrots
    - Francine: ice cream

    That night, Alfonse, Claire, and Dominic all came down with a wicked case of food poisoning. The others felt fine.

    a. Using only the Method of Agreement, how far can we narrow down the list of possible causes for the food poisoning?
    b. Using only the Method of Difference, how far can we narrow down the list of possible causes for the food poisoning?
    c. Using the Joint Method, we can identify the cause. What is it?

3. For each of the following, identify which of Mill's Methods is being used to draw the causal conclusion.
    a. A farmer noticed a marked increase in crop yields for the season. He started using a new and improved fertilizer that year, and the weather was particularly ideal—just enough rain and sunshine. Nevertheless, the increase was greater than could be explained by these factors. So he looked into it and discovered that his fields had been colonized by hedgehogs, who prey on the

kinds of insect pests that usually eat crops.

b. I've been looking for ways to improve the flavor of my vegan chili. I read on a website that adding soy sauce can help: it has lots of umami flavor, and that can help compensate for the lack of meat. So the other day, I made two batches of my chili, one using my usual recipe, and the other made exactly the same way, except for the addition of soy sauce. I invited a bunch of friends over for a blind taste test, and sure enough, the chili with the soy sauce was the overwhelming favorite!
c. The mere presence of guns in circulation can lead to higher murder rates. The data are clear on this. In countries with higher numbers of guns per capita, the murder rate is higher; in countries with lower numbers of guns per capita, the murder rate is correspondingly lower.
d. There's a simple way to end mass shootings: outlaw semiautomatic weapons. In 1996, Australia suffered the worst mass shooting episode in its history, when a man in Tasmania used two semiautomatic rifles to kill 35 people (and wound an additional 19). The Australian government responded by making such weapons illegal. There hasn't been a mass shooting in Australia since.
e. A pediatric oncologist was faced with a number of cases of childhood leukemia over a short period of time. Puzzled, he conducted thorough examinations of all the children, and also compared their living situations. He was surprised to discover that all of the children lived in houses that were located very close to high-voltage power lines. He concluded that exposure to electromagnetic fields causes cancer.
f. Many people are touting the benefits of the so-called "Mediterranean" diet because it apparently lowers the risk of heart disease. Residents of countries like Italy and Greece, for example, consume large amounts of vegetables and olive oil and suffer from heart problems at a much lower rate than Americans.
g. My daughter came down with what appeared to be a run-of-the-mill case of the flu: fever, chills, congestion, sore throat. But it was a little weird. She was also experiencing really intense headaches and an extreme sensitivity to light. Those symptoms struck me as atypical of mere influenza, so I took her to the doctor. It's a good thing I did! It turns out she had a case of bacterial meningitis, which is so serious that it can cause brain damage if not treated early. Luckily, we caught it in time and she's doing fine.

4.

CATEGORICAL LOGIC

Introduction

This chapter presents the first of two general methods used to evaluate deductive arguments. The method of the present chapter was developed by Aristotle nearly 2,500 years ago, and we'll refer to it simply as Aristotelian logic or at other times as categorical logic. We will look in greater detail at the second method, sentential logic, in later chapters.

Deductive Logics

Recall, deductive arguments are valid if and only if their premises guarantee their conclusions, and validity is determined entirely by the form of the argument. The two logics we study will have different ways of identifying the logical form of arguments, and different methods of testing those forms for validity. These are two of the things a deductive logic must do: specify precise criteria for determining logical form and develop a way of testing it for validity.

But before a logic can do those two things, there is a preliminary job: it must tame natural language. Real arguments that we care about evaluating are expressed in natural languages like English, Greek, etc. As we saw in our discussion of the logical fallacies, natural languages are unruly: they are filled with ambiguity and vagueness, and exhibit an overall lack of precision that makes it very difficult to conduct the kind of rigorous analysis necessary to determine whether or not an argument is valid. So before making that determination, a logic must do some tidying up; it must remove the imprecision inherent in natural language expressions of arguments and make them suitable for rigorous analysis. There are various approaches to this task. Aristotelian logic and sentential logic adopt two different strategies.

Aristotelian logic seeks to tame natural language by restricting itself to a well-behaved, precise portion of the language. It only evaluates arguments that are expressed within that precisely delimited subset of the language. Sentential logic achieves precision by eschewing natural language entirely: it constructs its own artificial language, and only evaluates arguments expressed in its terms. Thus, natural language arguments are evaluated in sentential logic by first translating the statements into this artificial language.

This strategy may seem overly restrictive: if we limit ourselves to arguments expressed in a limited vocabulary—and especially if we leave behind natural language—aren't we going to miss lots of arguments that we care about? The answer is no: these approaches are not nearly as restrictive as they might seem. We can translate back and forth between the special portion of language in Aristotelian logic and expressions in natural language that are outside its scope. Likewise, we can translate back and forth between the artificial language of sentential logic and natural language. The process of translating from the unruly bits of natural language into these more precise alternatives is what removes the ambiguity, vagueness, etc. that stand in the way of rigorous analysis and evaluation. So, part of the task of taming natural language is showing how one's alternative to it is nevertheless related to it—how it picks out the logically important features of natural language arguments while leaving behind their extraneous, recalcitrant bits.

These, then, are the three tasks that a deductive logic must accomplish:

1. Tame natural language.
2. Precisely define logical form.
3. Develop a way to test logical forms for validity.

The process for evaluating real arguments expressed in natural language is to render them precise and suitable for evaluation by translating them into the preferred vocabulary developed in step 1, then to identify and evaluate their forms according to the prescriptions of steps 2 and 3.

We now proceed to discuss Aristotelian logic, starting with its approach to taming natural language.

Classes and Categorical Propositions

For Aristotle, the fundamental logical unit is the class. Classes are just groups of things, or more formally sets of things—sets that we can pick out using language. The simplest way to identify a class is by using a plural noun—trees, clouds, asteroids, people—these are all classes. Names for classes can be grammatically more complex, too. We can modify the plural noun with an adjective: "rich people" picks out a class.

Prepositional phrases can further specify: "rich people from Italy" picks out a different class. The modifications can go on indefinitely: "rich people from Italy who made their fortunes in real estate and whose grandmothers were rumored to be secret lovers of Benito Mussolini" picks out yet another class—which is either very small, or possibly empty, I don't know. (Empty classes are just classes with no members; we'll talk more about them later.)

We will refer to names of classes as "class-terms," or just "terms" for short. Since for Aristotle the fundamental logical unit is the class, and since terms are the bits of language that pick out classes, Aristotle's logic is often referred to as a "term logic." This is in contrast to the logic we will study in the next chapter, sentential logic, so-called because it takes the fundamental logical unit to be the proposition (otherwise known as a "statement"), and sentences are the linguistic vehicle for picking those out.

Of course, Aristotelian logic must also deal with propositions—we're evaluating arguments here, and by definition those are just sets of propositions—but since classes are the fundamental logical unit, Aristotle restricts himself to a particular kind of proposition: categorical propositions.

"Category" is just a synonym of "class." Categorical propositions are propositions that make a claim about the relationship between two classes. This is the first step in taming natural language: Aristotelian logic will only evaluate arguments made up entirely of categorical propositions. We're limiting ourselves to a restricted portion of language—sentences expressing these kinds of propositions, which will feature two class terms—terms picking out the classes whose relationship is described in the categorical proposition. Soon, we will place further restrictions on the forms these sentences can take, but for now we will discuss categorical propositions generally.

Again, categorical propositions make an assertion about the relationship between two classes. There are three possibilities here:

1. **Whole Inclusion**: one class is contained entirely within the other.

   Example: Class 1 = people; Class 2 = bipeds. The first class is entirely contained in the second; every person is a biped.

2. **Partial Inclusion**: one class is partially contained within the other; the two classes have at least one member in common.

   Example: Class 1 = people; Class 2 = swimmers. Some people swim; some don't. Some swimmers are people; some aren't (e.g., fish). These two classes overlap, but not entirely.

3. **Exclusion**: the two classes don't have any members in common; they are exclusive.

   Example: Class 1 = people; Class 2 = birds. No people are birds; no birds are people. Batman notwithstanding (dude's not really a bat, and also, bats aren't birds; robins are birds, but again, Robin's not actually a bird, just a guy who dresses up like one).

Given these considerations, we can (more or less) formally define categorical propositions:

> A categorical proposition is a claim about the relationship between two classes—call them S and P—that either affirms or denies that S is wholly or partially included in P.[1]

1. Note that denying that S is even partially included in P is the same as affirming that S and P are exclusive.

Aristotle noted that, given this definition, there are four types of categorical proposition. We will discuss them in turn.

The Four Types of Categorical Proposition

TYPE 1: Universal affirmative (A)[2]

This type of proposition affirms the whole inclusion of the class S in the class P—it says that each member of S is also a member of P. The canonical expression of this proposition is a sentence of the form "All S are P."

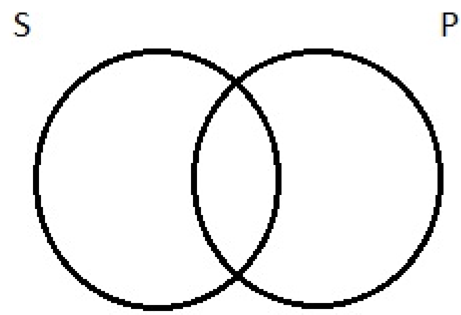
It is worth noting at this point why we chose "S" and "P" as the symbols for generic class terms. That's because the former is the grammatical subject (S) of the sentence, and the latter is the grammatical predicate (P). This pattern will hold for the other three types of categorical proposition.

Back to the universal affirmative: A proposition. It affirms whole inclusion. For example, the sentence "All men are mortals" expresses a proposition of this type: one that is true. "All men are Canadians" also expresses a universal affirmative proposition: one that is false.

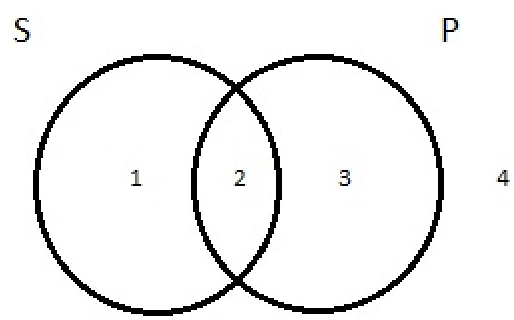
For the sake of concreteness, let's choose subject and predicate classes that we can use as go-to examples as we talk about each of the four types of categorical proposition. Let's let S = logicians and P = jerks. The A proposition featuring these two classes is expressed by "All logicians are jerks." (We'll remain agnostic about whether it's true or false.)

When it comes time to test arguments for validity—the last step in the process we've just begun—it will be convenient for us to represent the four types of categorical propositions pictorially. The basic form of the pictures will be two overlapping circles, with the left-hand circle representing the subject class and the right-hand circle representing the predicate class. Like this:

2. Since "universal affirmative"—along with the names of the other three types of categorical proposition—is a bit of a mouthful, we will follow custom and assign the four categoricals (shorthand for "categorical propositions") single-letter nicknames. The universal affirmative is the A proposition.



To depict the four types of categorical propositions, we'll modify this basic two-circle diagram by shading in parts of it or making marks inside the circles. Before we get to the specific depiction of the A proposition, though, let's talk about what the basic two-circle diagram does. It divides the universe into four regions, to which we can assign numbers like this:



Let's talk about what's inside each of the four regions if we take S to be the class of logicians and P to be the class of jerks.

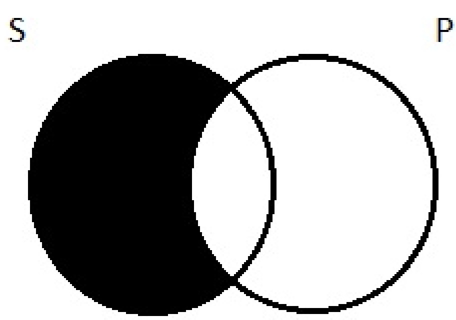
Region 1 is the portion of the S circle that doesn't overlap with the P circle. These are things in the subject class but outside the predicate class; they are logicians who aren't jerks. I never met Aristotle, but there's no evidence in the historical record to indicate that he was anything but a gentleman. So Aristotle is one of the residents of region 1—a logician who's not a jerk.

Region 2 is the area of overlap between the subject and predicate classes; its residents are members of both. So here we have the logicians who are also jerks. Gottlob Frege, a 19th century German logician, is the most important innovator in the history of logic other than Aristotle. Also, it turns out, he was a huge jerk. He was a big time anti-Semite. So Frege lives in region 2; he's both a logician and a jerk.

Region 3 is the portion of the P circle that doesn't overlap with S. These are members of the predicate class—jerks, in our example—who are not members of the subject class—not logicians. This is where the non-logician jerks live. Martin Shkreli is a resident of region 3.[3] The guy is clearly a jerk—and just as clearly, not a logician.

Region 4 is—everything else. It's all the things that are outside both the subject and predicate classes—things that are neither logicians nor jerks. You know who seems nice, but isn't a logician? Beyoncé. She lives in region 4. But so do lots and lots and lots of other things: the planet Jupiter is neither a logician nor a jerk; it's in there with Beyoncé, too. As is the left front tire of my wife's car. And the second-smallest brick in the Great Wall of China. And so on.

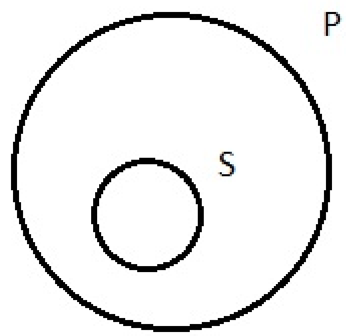
So much for the blank two-circle diagram and how it carves up the universe. What we want to figure out is *how to alter that diagram* so that we end up with a picture of the universal affirmative proposition. Our particular example of an A proposition is that all logicians are jerks. How do we draw a picture of that, using the two circles as our starting point? Well, think about it this way: when we say all logicians are jerks, what we're really saying is that a certain kind of thing doesn't exist; there's no such thing as a non-jerky logician. In other words, despite what I said above about Aristotle, region 1 is empty, according to this proposition (which, again, may or may not be true; it doesn't matter whether it's true or not; we're just trying to figure out how to draw a picture that captures the claim it makes). To depict emptiness, we will adopt the convention of shading in the relevant region(s) of the diagram. So our picture of the universal affirmative looks like this:



All S are P means that you won't find any members of S that are outside the P circle (no logicians who aren't jerks). The place in the diagram where they might've been is blotted out to indicate its emptiness. The only portion of S that remains as a viable space is inside the P circle, in what we called region 2 (the logicians you do find will all be jerks).

A reasonable question could be raised at this point: why did we draw the universal affirmative that way, instead of another, possibly more intuitive way? A propositions affirm whole inclusion—that S is entirely contained within P. Isn't the obvious way to depict that state of affairs more like this:

3. Martin Shkreli was widely criticized in 2015 for his decision to dramatically raise the price of Daraprim (a drug commonly used during HIV treatment to help prevent infections). The price went from $13.50 to $750 overnight, prompting medical groups to call the hike "unjustifiable."



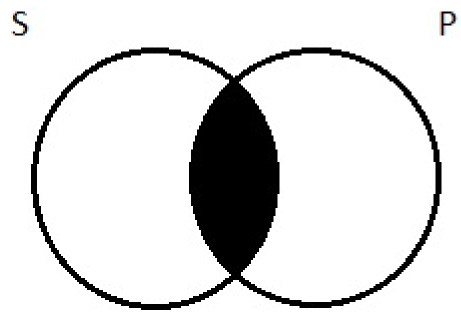
S entirely contained within P. Easy. Why bother with the overlapping circles and the shading?

There's nothing wrong with this alternative depiction of the universal affirmative; it captures the claim being made. We adopt the first alternative depiction for purely practical reasons: when it comes time to test arguments for validity, we're going to use these pictures, and our method will depend on our four types of categorical propositions all being depicted with the same basic two-overlapping-circle diagram, with shading and marks inside. These diagrams, as you may know, are called Venn diagrams. They are named after the 19th century English logician John Venn, who invented them specifically as an easier means of testing arguments for validity in Aristotelian logic (things were more unwieldy before Venn's innovation). It turns out Venn's method only works if we start with the overlapping circles for all four of the types of categorical proposition. So that's what we go with.

TYPE 2: Universal negative (E)

This type of proposition denies that S is even partially included in P. Put another way: it affirms that S and P are exclusive—that they have no members in common. The canonical expression of this proposition is a sentence of the form "No S are P." So, for example, the sentence "No dogs are cats" expresses a true universal negative proposition; the sentence "No animals are cats" expresses a false one.

Again, we want to think about how to depict this type of proposition using the standard two-circle Venn diagram. Think about the proposition that no logicians are jerks. How do we draw a picture of this claim? Well, as we said, E propositions tell us that the two classes don't have any members in common. The region of the two-circle diagram where there are members of both classes is the area of overlap in the picture (what we referred to as region 2 above). The universal negative proposition tells us that there's nothing in there. So if I claim that no logicians are jerks, I'm saying that, contrary to my claims above about the jerkiness of Gottlob Frege, no, there's no such thing as a logician-jerk. Region two is empty, and so we shade it out:



TYPE 3: Particular affirmative (I)

This type of proposition affirms that S is partially included in P. Its canonical expression is a sentence of the form "Some S are P." So, for example, "Some sailors are pirates" expresses a true particular affirmative proposition; "Some sumo wrestlers are pigeons" expresses a false one.

Before we talk about how to depict I propositions with a Venn diagram, we need to discuss the word "some." Remember, in Aristotelian logic we're taming natural language by restricting ourselves to a well-behaved portion of it—sentences expressing categorical propositions. We're proposing to use sentences with the word "some" in them. "Some," however, is not particularly well-behaved, and we're going to have to get it in line before we proceed.

Consider this utterance: "Some Republican voters are gun owners." This is true, and it communicates to the listener the fact that there's some overlap between the classes of Republican voters and gun owners. But it also communicates something more—namely, that some of those Republicans aren't gun owners. This is a fairly typical implicature: when we say that some are, we also communicate that some are not.

But there are times when we use "some" and don't implicate that some are not. Suppose you're talking to your mom, and you mention that you're reading a logic book. For some reason, your mom's always been curious about logic books, and asks you whether they're a good read. You respond, "Well, Mom, I can tell you this for sure: Some logic books are boring. You should see this book I'm reading now; it's a total snooze-fest!" In this case, you say that some logic books are boring based on your experience with this particular book, but you do not implicate that some logic books are not boring; for all you know, all logic books are boring—it's just impossible to write an exciting logic book. This is a perfectly legitimate use of the word "some," where all it means is that *there is at least one*: when you utter "some logic books are boring," all you communicate is that there is *at least one* boring logic book (this one, the one you're reading).

This is a bit of natural-language unruliness that we must deal with: sometimes when we use the word "some," we implicate that some are not; other times, we don't, only communicating that at least one is. When we use "some" in Aristotelian logic, we need to know precisely what's being said. So we choose:

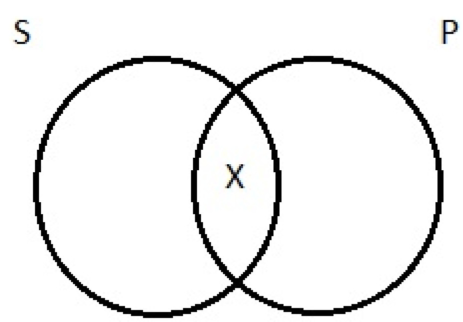
"Some" means: "there is at least one"

"Some S are P" tells us that those two classes have at least one member in common, and nothing more. "Some sailors are pirates" means that there's at least one sailor who's also a pirate, and that's it. There is no implication that some sailors are not pirates. All we mean to say is that at least one of them is, and for all we know, all of them are.[4]

This can confuse people, so it's worth repeating. Heck, let's indent it:

"Some" means "there is at least one," and that's it. It does not imply that some aren't.

With that out of the way, we can turn our attention to the Venn diagram for the particular affirmative. It makes the assertion that S and P have at least one member in common. Turning to our concrete example, the sentence "Some logicians are jerks" makes the claim that there is at least one logician who is a jerk.[5] How do we draw a picture of this? We need to indicate that there's at least one thing in the area of overlap between the two circles on the diagram—at least one thing inside of region 2. We do this by drawing an X:



TYPE 4: Particular negative (O)

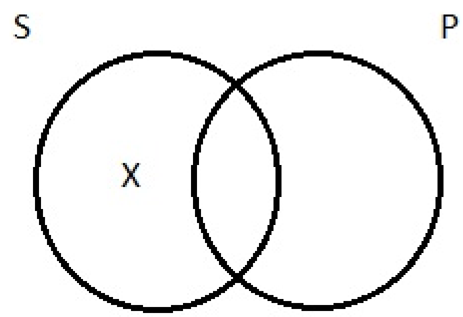
This type of proposition denies that S is wholly included in P. It claims that there is at least one member of S that is not a member of P. Given that "some" means "there is at least one," the canonical expression of this proposition is "Some S are not P"—there's at least one member of S that the two classes do not have in common. "Some sailors are not pirates" expresses a true particular negative proposition; "Some dogs are not animals" expresses a false one.

The Venn diagram for O propositions is simple. We need to indicate, on our picture, that there's at least one

4. The justification for this choice requires an argument, which I will not make here. The basic idea is that the "some aren't" bit that's often communicated is not part of the core meaning of "some"; it's an implicature, which is something that's (often, but not always) communicated over and above the core meaning.

5. In fact, this is true: Gottlob Frege was an anti-Semitic jerk.

thing that's inside of S, but outside of P. To depict the fact that some logicians are not jerks, we need to put Aristotle (again, not a jerk, I'm pretty sure) inside the S circle, but outside the P circle. As with the diagram for the I proposition, we indicate the existence of at least one thing by drawing an X in the appropriate place:



A Note on Terminology

It is commonly said that the four types of categorical propositions each have a quantity and a quality. There are two quantities: universal and particular. There are two qualities: affirmative and negative. There are four possible combinations of quantity and quality, hence, four types of categorical proposition.

The universal propositions—A and E, affirmative and negative—are so-called because they each make a claim about the entire subject class. If I claim that all hobos are whiskey drinkers, I've made an assertion that covers every single hobo, every member of that class. Similarly, if I claim that no chickens are race car drivers, I've made an assertion covering all the chickens—they all fail to drive race cars.

The particular propositions—I and O, affirmative and negative— on the other hand, do not make claims about every member of the subject class. "Some dinosaurs were herbivores" just makes the claim that there was at least one plant-eating dinosaur; we don't learn about all the dinosaurs.

Similar remarks apply to an O proposition like "Some dinosaurs were not carnivores." Remember, "some" just means "at least one."

The affirmative propositions—A and I, universal and particular—make affirmative claims about the relationship between two classes. A propositions affirm whole inclusion; I propositions affirm partial inclusion. Trivial fact: the Latin word meaning "I affirm" is affirmo; the A and the I in that word are where the one-letter nicknames for the universal and particular affirmatives come from.

The negative propositions—E and O, universal and particular—make negative claims about the relationship between two classes. E propositions deny even partial inclusion; O propositions deny whole inclusion. Trivial fact: the Latin word meaning "I deny" is nego; the E and the O in that word are where the one-letter nicknames for the universal and particular negatives come from.

Standard Form for Sentences Expressing Categorical Propositions

To tame natural language, Aristotelian logic limits itself to that portion of language that expresses categorical propositions. Above, we gave “canonical” sentences for each of the four types of categorical proposition: “All S are P” for the universal affirmative, “No S are P” for the universal negative, “Some S are P” for the particular affirmative, and “Some S are not P” for the particular negative. These are not the only ways of expressing these propositions in English, but we will restrict ourselves to these standard forms. That is, we will only evaluate arguments whose premises and conclusion are expressed with sentences with these canonical forms.

Generally speaking, here is the template for sentences qualifying as standard form:

[Quantifier] Subject Term <copula> (not) Predicate Term

Standard form sentences begin with a quantifier—a word that indicates the quantity of the categorical proposition expressed. **Restriction**: only sentences beginning with “All,” “No,” or “Some” qualify as standard form.

Subject and predicate terms pick out the two classes involved in the categorical proposition. **Restriction**: subject and predicate terms must be nouns or noun-phrases (nouns with modifiers) in order for a sentence to be in standard form.

The copula is a version of the verb “to be” (“are,” “is,” “were,” “will be,” etc.). **Degree of freedom**: it doesn’t matter which version of the copula occurs in the sentence; it may be any number or tense. For example, “Some sailors are pirates” and “Some sailors were pirates” both count as standard form.[6]

The word “not” occurs in the standard form expression of the particular negative, O proposition: “Some sailors are not pirates.” **Restriction**: the word “not” can only occur in sentences expressing O propositions; “not” appearing with any quantifier other than “some” is a deviation from standard form.

We now have a precise delimitation of the portion of natural language to which Aristotelian logic restricts itself: *only sentences in standard form*. But now a worry that we raised earlier becomes acute: if we can only evaluate arguments whose premises and conclusions are expressed with standard form sentences, aren’t we severely, perhaps ridiculously, constrained? Has anyone, ever, outside a logic book, expressed a real-life argument that way?

This is where translation comes in. Lots of sentences that are not in standard form can be translated into standard form sentences that have the same meaning. Aristotle himself believed that all propositions, no matter how apparently complex or divergent, could ultimately be analyzed as one of the four types of categorical proposition. Though this is, to put it mildly, not a widely held belief today, it still had an enormous influence in

---

6. Aristotelian logic is blind to tense: present, past, future, past perfect, future perfect, etc. are all the same. Sometimes the validity of an inference depends on tense. Aristotelian logic cannot make such judgments. This is one of the consequences of limiting ourselves to a simpler, more precise portion of natural language. There are more advanced logics that take verb tense into consideration (they’re unsurprisingly called “tense logics”), but that’s a topic for a different book.

the history of logic, since Aristotle's system was preeminent for more than 2,000 years. Over that time, logicians developed ever more elaborate procedures for analyzing a dizzying variety of non-standard form sentences as expressing one of the four types of categorical propositions, and translating them accordingly. An exhaustive survey of those inquiries would be beyond the scope of this book. It will be enough to look at a few simple examples to get an idea of how many apparently deviant expressions can be treated by Aristotelian logic. Our goal is simply to allay concerns that in restricting ourselves to standard form sentences we are severely limiting our logic's power to evaluate real-life arguments.

Let's consider a famous deductively valid argument, the one about Socrates: All men are mortal; Socrates is a man; therefore, Socrates is mortal. This argument has three propositions in it, but none of the three sentences expressing them are in standard form. The first sentence, "All men are mortal," may appear to fit the bill, but it has a subtle flaw: "mortal" is an adjective, not a noun. Class terms are required to be nouns or noun phrases. But this is an easy fix: add an "s" to the end and you get a plural noun. "All men are mortals" is in standard form; it expresses a universal affirmative, A proposition. This prescription applies generally. Predicate adjectives can be replaced with suitable noun phrases most easily by just inserting the generic noun "things": "Some men are handsome" becomes "Some men are handsome things"; "No priests are silly" becomes "No priests are silly things."

Back to the Socrates argument. The second premise is also problematic: "Socrates is a man." First of all, it doesn't have a quantifier. Second, its subject term, "Socrates," picks out an individual person; we're supposed to be dealing with classes here, right? Well, that's right, but it's not really a problem. We can just make the subject class a unit class—a class containing exactly one member, namely Socrates. Now we can understand the sentence as expressing the claim that the single member of that class is also a member of the class of men. That is, it's a universal affirmative—there's whole inclusion of the Socrates unit-class in the class of men. The sentence we need, then, starts with the quantifier "All," and to make the grammar work, we pick a plural noun to name the Socrates class: "All Socrateses are men." Is "Socrateses" the plural of "Socrates"? I can't think of anything better (we could try "All individuals identified as the one Socrates" to convey the same entity we want our term to pick out). Anyway, the point is, that word picks out a class that has exactly one member, Socrates. Sentences with singular subjects can be rendered as universals. If I had the sentence "Socrates is not alive," I could render it as a universal negative: "No Socrateses are living things."

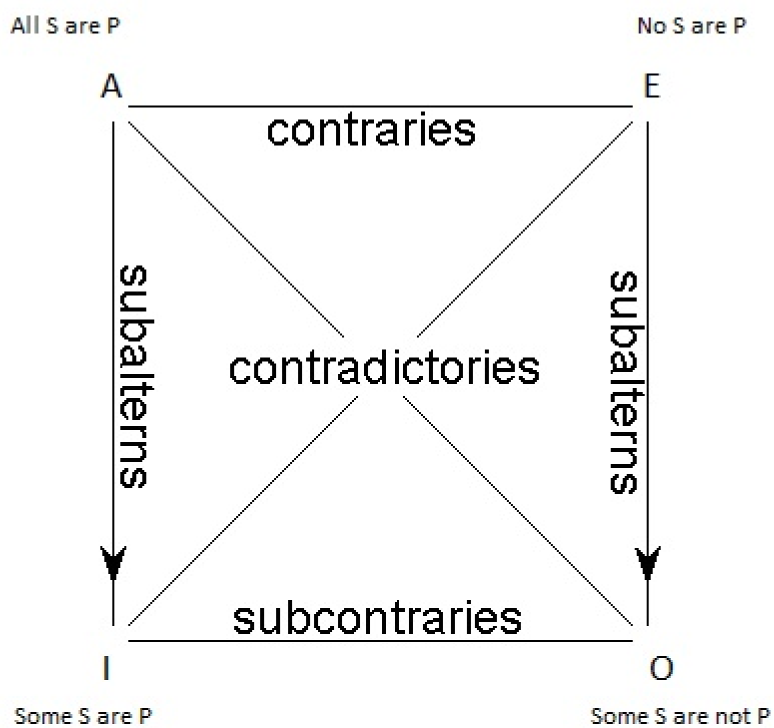
There are other things to consider. English comes with a variety of quantifier words: "each," "every," "any," and so on. Common sense tells us how to translate sentences featuring these into standard form: switch to the appropriate standard form quantifier—"All," "No," or "Some." "Every teacher is a hard worker" becomes "All teachers are hard workers," for example. Sometimes quantifier words are omitted, but it's clear from context what's going on. "Dogs are animals" means "All dogs are animals"; "People are waiting in line" can be rendered as "Some people are things that are waiting in line." Some sentences have a verb other than the copula. "Some people eat rabbit," for example, can be translated into "Some people are rabbit-eaters." Sometimes the word "not" appears in a sentence that has a quantifier other than "some." "Not all mammals are carnivores," for example, can be translated into "Some mammals are not carnivores."

The list goes on. As I said, centuries of work have been done on the task of translating sentences into standard form. We can stop here, I think, and simply accept that the restriction to standard form sentences does not seriously limit the arguments that Aristotelian logic can evaluate.

The Square of Opposition

Having established the boundaries of our domain of logically well-behaved natural language, we turn now to an investigation of the properties of its inhabitants. The four types of categoricals are related to one another in systematic ways; we will look at those relationships.

The relationships are inferential: we can often infer, for example, from the truth of one of the four categoricals, whether the other three are true or false. These inferential relationships among the four categorical propositions are summarized graphically in a diagram: The Square of Opposition. The diagram looks like this:



The four types of categorical propositions are arranged at the four corners of the square, and along the sides and diagonals are marked the relationships between pairs of them. We take these relationships up in turn.

Contradictories

Contradictory pairs of categorical propositions are at opposite corners from one another on the Square of Opposition. A and O propositions are contradictory; E and I propositions are contradictory. What it means for a pair of propositions to be contradictory is this: they have opposite truth-values; when one is true, the other must be false, and *vice versa*.

This is pretty intuitive. Consider an A proposition—all sailors are pirates. Suppose I make that claim. How do you contradict me? How do you prove I'm wrong? "My brother's in the Navy," you might protest. "He's a sailor, but he's not a pirate." That would do the trick. The way you contradict a universal affirmative claim—a claim that all S are P—is by showing that there's at least one S (a sailor in this case, your brother) who's not a P (not a pirate, as your brother is not). At least one S that's not a P—that's just the particular negative, O proposition, that some S are not P. (Remember: "some" means "there is at least one.") A and O propositions make opposite, contradictory claims. If it's false that all sailors are pirates, then it must be true that some of them aren't; that's just how you show it's false. Likewise, if it's true that all dogs are animals (it is), then it must be false that some of them are not (you're not going to find even one dog that's not an animal). A and O propositions have opposite truth values.

Likewise for E and I propositions. If I claim that no saints are priests, and you want to contradict me, what you need to do is come up with a saint who was a priest. It's not hard: Saint Thomas Aquinas (who was the most prominent medieval interpreter of Aristotle and a terrific philosopher in his own right) was a priest. So, to contradict the universal negative claim—that no S are P—you need to show that there's at least one S (a saint in this case, Thomas Aquinas) who is in fact a P (a priest, as Aquinas was). At least one S that is a P—that's just the particular affirmative, I proposition, that some S are P. (Again, "some" means "there is at least one.") E and I propositions make opposite, contradictory claims. If it's false that no saints are priests, it must be true that some of them are; that's just how you show it's false. Likewise, if it's true that no cats are dogs (it is), then it must be false that some of them are (you're not going to find even one cat that's a dog). E and I propositions have opposite truth values.

Contraries

The two universal propositions—A and E, along the top of the square—are a *contrary* pair. This is a slightly weaker form of opposition than being *contradictory*. Being contrary means that they can't both be true, but they *could* both be false—though they needn't both be false; one could be true and the other false.

Again, this is intuitive. Suppose I claim the universal affirmative, "All dogs go to heaven," and you claim the corresponding universal negative, "No dogs go to heaven." (Those sentences aren't in standard form, but the translation is easy.) Obvious observation: we can't both be right; that is, both claims can't be true. On the other hand, we could both be wrong. Suppose getting into heaven, for dogs, is the way they say it is for people: if you're good and stuff, then you get in, but if you're bad, oh boy—it's the Other Place for you. In that case, both of our claims are false: some dogs (the good ones) go to heaven, but some dogs (the bad ones, the ones who bite kids, maybe) don't. But that picture might be wrong, too. I could be right and you could be wrong: God loves all dogs equally and they get a free pass into heaven. Or, I could be wrong and you could be right: God hates dogs and doesn't let any of them in; or maybe there is no heaven at all, and so nobody goes there, dogs included.

Subcontraries

Along the bottom of the square, we have the two particular propositions—I and O—and they are said to be *subcontraries*. This means they can't both be false, but they *could* both be true—though they needn't be; one could be true and the other false.

It's easy to see how both I and O could be true. As a matter of fact, some sailors are pirates. That's true. Also, as a matter of fact, some of them are not. It's also easy to see how one of the particular propositions could be true and the other false, provided we keep in mind that "some" just means "there is at least one." It's true that some dogs are mammals—that is, there is at least one dog that's a mammal—so that I proposition is true. In fact, all of them are—the A proposition is true as well.

Which means, since A and O are contradictories, that the corresponding O proposition—that some dogs are not mammals—must be false. Likewise, it's true that some women are not (Catholic) priests (at least one woman isn't a priest), and it's false that some women are priests (the Church doesn't allow it). So O can be true while I is false.

It's a bit harder to see why both particular propositions can't be false. Why can't "Some surfers are priests" and "Some surfers are not priests" both be false? It's not immediately obvious. But think it through: if the I (some surfers are priests) is false, that means the E (no surfers are priests) must be true, since I and E are contradictory. If the O (some surfers are not priests) is false, that means the A (all surfers are priests) must be true, since O and A are contradictory. That is to say, if I and O were both false, then the corresponding A and E propositions would both have to be true. But, as we've seen already, this is (obviously) impossible: if I claim that all surfers are priests and you claim that none of them are, we can't both be right.

Subalterns

The particular propositions at the bottom of the table—I and O—are *subalterns* of the universal propositions directly above them—A and E, respectively.[7] This means that the pairs have the following relationship: if the universal proposition is true, then the particular proposition (its subaltern) must also be true. That is, if an A proposition is true, its corresponding I proposition must also be true; if an E proposition is true, its corresponding O proposition must also be true.

This is intuitive if we keep in mind, as always, that "some" means "there is at least one." Suppose the A proposition that all whales are mammals is true (it is): then the corresponding I proposition, that some whales are mammals, must also be true. Again, "some whales are mammals" just means "at least one whale is a mammal"; if all of them are, then at least one of them is! Similarly, on the negative side of the square, if it's

7. And the universal propositions are called superalterns.

true that no priests are women (universal negative, E), then it's got to be true that some priests are not women (particular negative, O)—that at least one priest is not a woman. If none of them are, then at least one isn't!

Notice that these relationships are depicted in a slightly different way from the others on the Square of Opposition: there's an arrow pointing toward the bottom. This is because the relationship is not symmetrical. If the proposition on top is true, then the one on the bottom must also be true, but the reverse is not the case. If an I proposition is true—some sailors are pirates—it doesn't follow that the corresponding A proposition—that all sailors are pirates—is true. Likewise, the truth of an O proposition—some surfers are not priests—does not guarantee the truth of the corresponding E proposition—that no surfers are priests.

Truth, as it were, travels down the side of the square. Falsehood does not: if the universal proposition is false, that doesn't tell us anything about the truth or falsehood of the corresponding particular. You could have a false A proposition—all men are priests—with a true corresponding I—some men are priests. But you could also have a false A proposition—all cats are dogs—whose corresponding I—some cats are dogs—is also false. Likewise, you could have a false E proposition—no men are priests—with a true corresponding O—some men are not priests. But you could also have a false E proposition—no whales are mammals—whose corresponding O—some whales are not mammals—is also false.

Falsehood doesn't travel down the side of the square, but it does travel up. That is, if a particular proposition—I or O—is false, then its corresponding universal proposition—A or E, respectively—must also be false. Think about it in the abstract: if it's false that some S are P, that means that there's not even one S that's also a P. In that case, there's no way *all* the Ss are Ps! False I, false A. Likewise on the negative side: if it's false that some S are not P, that means you won't find even one S that's not a P, which is to say *all* the Ss are Ps; in that case, it's false that no S are P (A and E are contraries). False O, false E.

Inferences

Given information about the truth or falsity of a categorical proposition, we can use the relationships summed up in the Square of Opposition to make inferences about the truth values of the other three types of categorical proposition.

Here's what I mean. Suppose a universal affirmative proposition—an A proposition—is true. What are the truth values of the corresponding E, I, and O propositions? (By "corresponding," I mean propositions with the same subject and predicate classes.) The square can help us answer these questions. First of all, A is in the opposite corner from O—they're contradictory. That means A and O have to have opposite truth values. Well, if A is true, as we're supposing, then the corresponding O proposition has to be false. Also, A and E are contraries. That means that they can't both be true. Well, we're supposing that the A is true, so the corresponding E must be false. What about the I proposition? Three ways to attack this one, and they all agree that the I must be true: (1) I is the subaltern of A, so if A is true, then I must be true as well; (2) I is the contradictory of E, and we've already determined that E must be false, so I must be true; (3) I and O are subcontraries, meaning they can't both be false, and since we've already determined that O is false, it follows that I must be true.

Summing up: if an A proposition is true, the corresponding E is false, I is true, and O is false.

Let's try another one: suppose a universal negative, E proposition, is true. What about the corresponding A, I, and O propositions? Well, again, A and E are contraries—can't both be true—so A must be false. I is the contradictory of E, so it must be false—the opposite of I's truth value. And since O is subaltern to E, it must be true because E is.

If an E proposition is true, the corresponding A is false, I is false, and O is true.

Another. Suppose a particular affirmative, I proposition, is true. What about the other three? Well, E is its contradictory, so it must be false. And if some S are P, that means some of them aren't— so the O is also true. And since A is the contradictory of O…WAIT JUST A MINUTE! Go back and read that again. Do you see what happened? "And if some S are P, that means some of them aren't…" No it doesn't! Remember, "some" means "there is at least one." If some S are P, that just means at least one S is a P—and for all we know, *all* of them might be; then again, maybe not. I and O are subcontraries: they can't both be false, they *could* both be true, and one could be true and the other false. Knowing that I is true tells us nothing about the truth value of the corresponding O or the corresponding A. That some are, meaning at least one is, leaves open the possibility that all of them are, but then again, maybe not. The fact is, based on the supposition that an I is true, we can only know the truth value of the corresponding E for sure.

If an I proposition is true, then the corresponding E is false, and A and O are of unknown truth value.

Exercises

1. Suppose an O proposition is true. What are the truth values of the corresponding A, E, and I propositions, according to the Square of Opposition?
2. Suppose an A proposition is false. What are the truth values of the corresponding E, I, and O propositions, according to the Square of Opposition?
3. Suppose an E proposition is false. What are the truth values of the corresponding A, I, and O propositions, according to the Square of Opposition?
4. Suppose an I proposition is false. What are the truth values of the corresponding A, E, and O propositions, according to the Square of Opposition?
5. Suppose an O proposition is false. What are the truth values of the corresponding A, E, and I propositions, according to the Square of Opposition?

Operations on Categorical Sentences

We continue our exploration of the portion of natural language to which Aristotle's logic restricts itself—the standard form sentences expressing categorical propositions. To familiarize ourselves more intimately with these, we will look at how they respond when we perform various operations on them, when we manipulate them in various ways.

We will examine three operations: **conversion**, **obversion**, and **contraposition**. Each of these alters the standard form sentences in some way. The question we will ask is whether the new sentence that results from the manipulation is equivalent to the original sentence; that is, does the new sentence express the same proposition as the original?

Conversion

Performing conversion on a categorical sentence involves changing the order of the subject and predicate terms. The result of this operation is a new sentence, which is said to be the *converse* of the original sentence. Our question is: when does performing conversion produce an equivalent new sentence, a converse that expresses the same proposition as the converted original? We will look at all four types of standard form sentence, answering the question for each.

Let's perform conversion on a sentence expressing a universal affirmative, A proposition, and see what happens. "All dogs are animals" is such a sentence. Conversion switches the subject and predicate terms, so the converse sentence is "All animals are dogs." Does the converse express the same proposition as the original? Are they equivalent? Heck, no! The original sentence expresses the true proposition that all dogs are animals; the converse expresses the utterly false proposition that all animals are dogs. Converting an A sentence produces a new sentence that is not equivalent to the original.

This means that the effect on truth value, in the abstract, of converting A sentences, is unpredictable. Sometimes, as with "All dogs are animals," conversion will lead you from a truth to a falsehood. Other times, it may lead from truth to truth: "All bachelors are unmarried men" and "All unmarried men are bachelors" express different propositions, but both of them are true (because it so happens that, by definition, a bachelor is just an unmarried man). Conversion of an A could also lead from falsehood to falsehood, as with the transition from "All dogs are bats" to "All bats are dogs." And it could lead from falsehood to truth: just reverse the order of the first conversion we looked at, from "All animals are dogs" to "All dogs are animals." Again, the point here is that:

> Because conversion of A sentences produces a converse that expresses a different proposition than the original, we cannot know what the effect of the conversion will be on truth value.

How about conversion of sentences expressing universal negative, E propositions? "No dogs are cats" is such a sentence. Its converse would then be "No cats are dogs." Are they equivalent? Yes, of course. Remember, an E proposition denies even partial inclusion; it makes the claim that the two classes involved don't have any members in common. It doesn't matter which of the two classes is listed first in the sentence expressing that proposition—you still get the assertion that the two classes are exclusive. This is true of E sentences generally:

> Performing conversion on them always produces a new sentence that is equivalent to the original.

It is also true of sentences expressing particular affirmative, I propositions. "Some sailors are pirates," after conversion, becomes "Some pirates are sailors." These express the same proposition: they make the claim that the two classes have at least one member in common—there is at least one thing that is both a sailor and a pirate. Again, it doesn't matter what order you put the class terms in; I sentences express the assertion that there's overlap between the two classes.

An I sentence and its converse are always equivalent.

The same cannot be said of sentences expressing particular negative, O propositions. Consider "Some men are not priests." That expresses a true proposition. But its converse, "Some priests are not men" expresses a different proposition; we know it's a different proposition because it's false.

That is all we need to show that an operation does not produce equivalent sentences: one counterexample. As above with A sentences, this means that:

The effect on truth value of converting O sentences is unpredictable.

It can take us from truth to falsehood, as in this example, or from truth to truth, falsehood to falsehood, falsehood to truth. In the abstract, we cannot know the effect on truth of converting O sentences, since the converse expresses a different proposition from the original.

Summary for Conversion: For E and I, converses are equivalent; for A and O, converses are not.

Obversion

Before we talk about our next operation, obversion, we need to introduce a new concept: class complements. The *complement* of a class, call it S, is another class which contains all the things that are not members of S. So, for example, the complement of the class of trees is just all the things that aren't trees. The easiest way to name class complements is just to stick the prefix "non" in front of the original class name. So the complement of trees is non-trees.

Be careful: it may be tempting, for example, to say that the complement of Republicans is Democrats. But that's not right. The complement of Republicans is a much bigger class, containing *all* the non-Republicans: not just Democrats, but Communists and Libertarians and Independents and Greens; oh, and a bunch of other things, too—like the planet Jupiter (not a Republican), my left pinkie toe, the Great Wall of China, etc.

As a matter of notational convention, if we use a capital letter like S to refer to a class, we will denote the complement of that class as ~ S, which we'll read as "*tilde-S*."

Back to obversion. Here's how this operation works: first, you change the quality of the sentence (from affirmative to negative, or vice versa); then, you replace the predicate with its complement. The result of performing obversion on a sentence is called the *obverse* of the original.

It turns out that performing obversion on a sentence always produces a new sentence that's equivalent to it; a sentence and its obverse always express the same proposition. That means they share a truth value: if a sentence is true, so is its obverse; if a sentence is false, its obverse is false, too. We can see that this is so by looking at the result of performing obversion on each of the four types of standard form sentences.

We'll start with A sentences. Consider "All ducks are swimmers." To perform obversion on this sentence, we first change its quality. This is a universal affirmative. Its quality is affirmative. So we change that to negative, keeping the quantity (universal) the same. Our new sentence is going to be a universal negative, E sentence—something of the form No S are P. Next, we replace the predicate with its complement. The predicate of the sentence is "swimmers." What's the complement of that class? All the things that aren't swimmers: non-swimmers. So the obverse of the original A sentence is this: "No ducks are non-swimmers."

Now, are these two sentences equivalent? Yes. "All ducks are swimmers" expresses the universal affirmative proposition, asserting that the class of ducks is entirely contained in the class of swimmers. That is to say, any duck you find will also be in the swimmer class. Another way of putting it: you won't find any ducks who *aren't* in the class of swimmers. In other words, no ducks fail to be swimmers. Or: "No ducks are non-swimmers." The A sentence and its obverse are equivalent; they express the same proposition, make the same claim about the relationship between the class of ducks and the class of swimmers.

Let's try obversion on a universal negative, E sentence. "No women are priests" is one. First, we change its quality from negative to affirmative: it becomes a universal affirmative, A sentence—something of the form All S are P. Next, we replace its predicate, "priests," with its complement, "non-priests." The result: "All women are non-priests." Is that equivalent to the original? It tells us that all women are outside the class of priests. In other words, none of them are priests. That is, "No women are priests." Yes, both the original sentence and its obverse tell us that the classes of women and priests are exclusive.

Next, the particular affirmative—an I sentence like "Some politicians are Democrats." OK. First, change the quality—from affirmative to negative. Our obverse will be a particular negative, O sentence—something of the form Some S are not P. Now, replace "Democrats" with "non-Democrats," stick it in the predicate slot, and we get "Some politicians are not non-Democrats." Well, that's not exactly grammatically elegant, but the meaning is clear: not being a non-Democrat is just being a Democrat. This says the same thing as the original, namely that some politicians are Democrats.

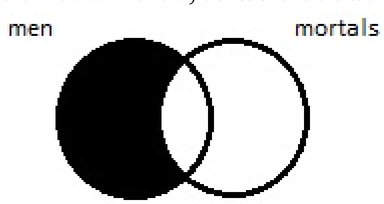
Finally, particular negative, O. We'll try "Some plants are not flowers." Changing from negative to affirmative means our obverse will be an I—Some S are P. We replace "flowers' with "non-flowers" and get "Some plants are non-flowers." We went from "Some plants are not flowers" to "Some plants are non-flowers." Obviously, those are equivalent.

Summary for obversion: obverses are equivalent for A, E, I, and O.

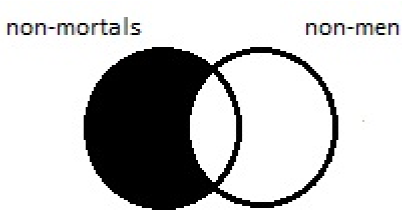
Contraposition

Our last operation is contraposition. Unlike obversion, and like conversion, it doesn't involve changing the type (A, E, I, O) of the sentence we're operating on. Rather, again, like conversion, we just manipulate the subject and predicate. Here's how: replace the subject with the complement of the predicate and replace the predicate with the complement of the subject. The result of performing contraposition on a sentence is called its *contrapositive*.

Let's perform contraposition on an A sentence: "All men are mortals." To form its contrapositive, we put the complement of the predicate—non-mortals—into subject position and the complement of the subject—non-men—into predicate position: "All non-mortals are non-men." The question, as always: are these sentences equivalent? This one's a bit hard to see. Let's use Venn diagrams to help us think it through. First, we know what the diagram for "All men are mortals" looks like; that sentence claims that there's no such thing as a man who's not a mortal, so we blot out the portion of the "men" circle that's not inside the "mortals" circle:

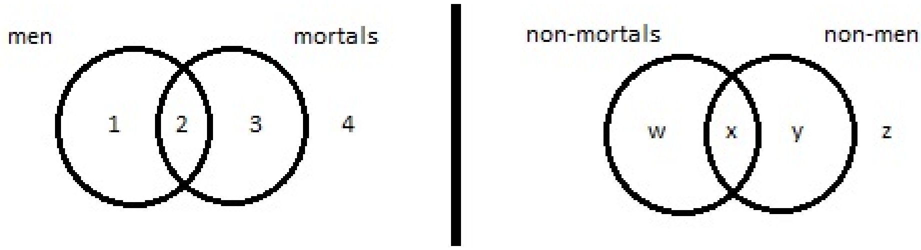


Next, let's think through how we would diagram "All non-mortals are non-men." If we change our circles to "non-men" and "non-mortals," respectively, it's easy; when you're diagramming an A proposition, you just blot out the part of the left-hand (subject) circle that doesn't overlap with the right-hand (predicate) circle. There's no such thing as non-mortals who aren't non-men:



But how do we compare this diagram with the one for "All men are mortals" to see if they express the same proposition? We need to know that the two would give us the same picture if the circles were labeled the same.

Let's compare the unshaded diagrams where the circles are "men" and "mortals," on the one hand, and "non-men" and "non-mortals" on the other:



When we depict “All men are mortals,” we blot out region 1 of the left-hand diagram. When we depict its contrapositive, “All non-mortals are non-men,” we blot out region w of the right-hand diagram. We want to know whether these two sentences are equivalent. They are, provided that blotting out region 1 and blotting out region w amount to the same thing. Do they? That is, do regions 1 and w contain the same objects?

Let’s think this through, starting with region z. What’s in there? Those are the things that are outside both the non-mortal and non-men circles; that is, they’re *not* non-mortals and they’re *not* non-men. So they’re mortals and men, right? Things that are both mortals and men: on the left-hand diagram, that’s the overlap between the circles. Region z and region 2 contain the same things.

How about region y? Those things are non-men, but they’re outside the non-mortals circle, making them mortals. Mortals who aren’t men: they live in region 3 in the left-hand diagram. Regions y and 3 contain the same things. Region x has things that are both non-men and non-mortals; that is, they’re outside both the mortal and men circles on the left. Regions x and 4 contain the same things.

And region w? Outside the non-men circle, so they’re men. Inside the non-mortals circle, so they’re not mortals. Men that aren’t mortals: that’s region 1 on the left. Regions w and 1 contain the same things. And that means that blotting out region w and blotting out region 1 amount to the same thing; both are ways of ruling out the existence of the same group of objects, the men who aren’t mortals—or, as it turns out, the non-mortals who aren’t non-men. Same thing.

Picking the main thread back up, what all this shows is that when we perform contraposition on universal affirmative, A sentences, we end up with new sentences that express the same proposition.

An A sentence and its contrapositive are equivalent.

We still have to ask the same question about E, I, and O sentences.

Consider a universal negative (E): “No skydivers are cowards.” This is surely true; it takes bravery to jump out of a plane (I wouldn’t do it). To get the contrapositive, we replace the subject, skydivers, with the complement of the predicate, non-cowards, and we replace the predicate, cowards, with the complement of the subject, non-skydivers. The result is “No non-cowards are non-skydivers.” That’s false. You know who was a non-coward? Martin Luther King, Jr. The Reverend King was a courageous advocate for racial equality up to the very last day of his life. But, not a skydiver. The contrapositive claims there’s no such thing as a non-coward who doesn’t sky-dive. But that isn’t so: MLK is a counterexample. In general, when you perform contraposition on an E sentence, you end up with a new sentence that expresses a different proposition. And as was the case with A and O sentences being converted, this has unpredictable effects on truth value. You may move from truth to falsehood, as in this case, or from truth to truth, falsehood to falsehood, falsehood to truth.

Contraposition changes the proposition expressed by E sentences, so you can't know the resulting truth value.

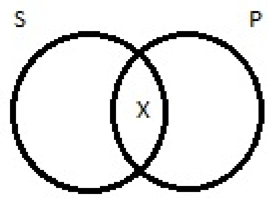
Next, consider particular negative (O) sentences. These are pretty easy. "Some men are not priests" is a good go-to example. Performing contraposition, we get "Some non-priests are not non-men." Things that are not non-men—those are just men. So the claim being made by the contrapositive is that some non-priests are men. That is, there's at least one thing that's both a non-priest and a man; or, there's at least one man who's not a priest. I know a way to say that: "Some men are not priests." The O sentence and its contrapositive make the same claim.

Contraposition performed on particular negatives gives you a new sentence that is equivalent to the original.

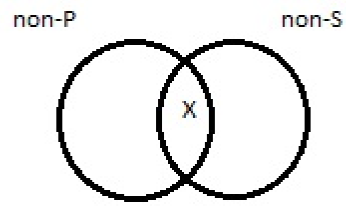
Finally, particular affirmatives—I sentences. "Some men are priests" is true. So is its contrapositive: "Some non-priests are non-men" (there's at least one: my mom is not a man, nor was she ever a priest). So contraposition performed on an I works? That is, it gives you an equivalent sentence? Not necessarily. The two sentences might both be true, but they could be expressing two different true propositions. As a matter of fact, they are.

When you contrapose an I sentence, the result is a new sentence that is not equivalent.

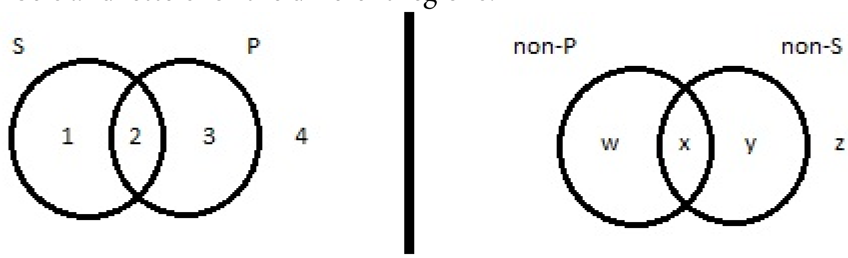
To see why, we'll return to Venn diagrams. Generically speaking, an I proposition's diagram has an X in the area of overlap between the two circles. For a sentence of the form Some S are P, we would draw this:



There is at least one thing (the X) that is both S and P. For the contrapositive, we draw this:



There is at least one thing that is both non-P and non-S. The question is, does drawing an X in those two regions of overlap amount to the same thing? Let's put the diagrams side by side, without the Xs, but with numbers and letters for the different regions:



We went through this above when we were discussing the effects of contraposition on A propositions. Regions 1 and w contain the same things, as do regions 3 and y. But regions 2 and 4 don't line up with regions x and z, respectively. Rather, they're reversed: region 2 has the same objects as region z, and region 4 has the same objects as region x.

When we draw the picture of the straight-up I sentence, we put an X in region 2; when we draw the picture of its contrapositive, we put an X in region x. But region 2 and region x aren't the same. So the I sentence and its contrapositive, in general, are not equivalent. Performing contraposition on an I sentence changes the proposition expressed, with unpredictable effects on truth value.

We can prove it with a concrete example. Let our starting I sentence be "Some Catholics are non-Popes." That's certainly true (again, my mom: Catholic, but not Pope). The contrapositive would be "Some Popes are non-Catholics" (the complement of non-Popes is just Popes). But that's false. Being Catholic is a prerequisite for the Papacy. An I sentence and its contrapositive make different claims.

Exercises

1. Perform conversion on the following and write down the converse. Is it equivalent to the original sentence?
    a. Some surfers are not priests.

   b. All Canadians are bodybuilders.
   c. No Mexicans are fishermen.
   d. Some Nazis are florists.
2. Perform obversion on the following and write down the obverse. Is it equivalent to the original sentence?
   a. No people are lizards.
   b. Some politicians are criminals.
   c. Some birds are not animals.
   d. All Democrats are samurais.
3. Perform contraposition on the following and write down the contrapositive. Is it equivalent to the original sentence?
   a. All Philistines are Syrians.
   b. No Africans are Europeans.
   c. Some Americans are Irishmen.
   d. Some Swiss are not Catholics.

Inferences

Earlier, we discussed how we could make inferences about the truth values of categoricals using the information encoded in the Square of Opposition. For example, given the supposition that an A sentence expresses a true proposition, we can infer that the corresponding E sentence expresses a falsehood (since A and E are contraries, which can't both be true), that the corresponding I sentence expresses a truth (since I is the subaltern of A, which means A's truth guarantees that of I), and that the corresponding O sentence expresses a falsehood (since A and O are contradictories, which must have opposite truth values).

The key word in that paragraph is "corresponding." The Square of Opposition tells us about the relationships among categoricals that *correspond*—which means they have the same subjects and predicates. If "All S are P" is true, then "No S are P" must be false, per the square, since these two sentences have the same subject (S) and predicate (P). The square *cannot* license such inferences when the subjects and predicates do not correspond. The supposition that "All S are P" is true tells me nothing at all about the truth value of "Some A are B." The subjects and predicates are different; we're dealing with two different classes.

There are occasions, however, when subjects and predicates do not correspond, but we can nevertheless make inferences about the truth values of categoricals based on information about others. In such cases, we need to combine our knowledge of the relationships depicted in the Square of Opposition with our recently acquired knowledge about the circumstances in which conversion, obversion, and contraposition provide us with equivalent sentences.

Here is a simple example. Suppose that a sentence of the form "No S are P" expresses a truth (never mind what "S" and "P" stand for; we're thinking in the abstract here). Given that information, what can we say about a sentence of the form "Some P are S"? Well, the first is an E and the second is an I. According to the Square of Opposition, E and I are a contradictory pair, so they must have opposite truth values. But remember, the relationships in the square only hold for *corresponding* sentences. "No S are P" and "Some P are S" do not correspond; their subject and predicate class terms are in different spots. The square tells us that the I sentence corresponding to "No S are P"—namely, "Some S are P"—must have the opposite truth value. We've presumed that the E sentence is true, so "Some S are P" expresses a falsehood, according to the square. But we wanted to know the truth value of "Some P are S," the sentence with the subject and predicate terms switched. Well, switched subject and predicate terms—that's just the *converse* of "Some S are P." And we know from our investigations that performing conversion on an I sentence always gives you another I sentence that's equivalent to the first; that is, it expresses the same proposition, so it's true or false in all the same circumstances as the original. That means "Some P are S" must express a falsehood, just like its converse.

Here's how to think about the inference we just made. We were given the fact that "No S are P" is true. We wanted to know the truth value of "Some P are S."[8] We can't compare these two directly using the Square of Opposition because they don't correspond: different subject and predicate. But we know that the converse of the our target sentence—"Some S are P"—*does* correspond, so according to the Square, it must be false (since it's contradictory to "No S are P"). And, since conversion on I sentences yields equivalent results, "Some P are S" has the same truth value as "Some S are P," so our target sentence must also be false.

This is the general pattern for these sorts of multi-step inferences. You're given information about a particular categorical claim's truth value, then asked to evaluate some other claim for truth or falsity. They may not correspond, so the first stage of your deliberations involves getting them to correspond—making the subject and predicate terms line up. You do this by performing conversion, obversion, and contraposition as needed, but only when those operations produce equivalent results: you only use conversion on E and I sentences; you only use contraposition on A and O sentences; and since obversion always yields an equivalent sentence, you can use it whenever you want. Then, once you've achieved correspondence, you can consult the Square of Opposition and complete the inference.

Another example can help illustrate the method. Suppose we're told that some sentence "All S are P" is true. What about the sentence "No ~ S are ~ P"? (Remember, when we put the tildes in front of the letters, we're referring to the complements of these classes.) First, we notice that the subject and predicate terms don't correspond. The A sentence has "S" in subject position and "P" in predicate position, while the target E sentence has ~ S and ~ P in those slots. We can see this misalignment clearly (and also set ourselves up to think through the remaining steps in the inference more easily) if we write the sentences out, one above the other (noting in brackets what we know about their truth values):

---

8. We're getting a little sloppy here. Technically, it's propositions, not sentences, that are true or false. Further complication: we're not even talking about actual sentences here, but generic sentence patterns, with placeholder letters "S" and "P" standing in for actual class terms. Can those sorts of things be true or false? Ugh. Let's just agree not to be fussy and not to worry about it. We all understand what's going on.

All S are P [T]

No ~ S are ~ P [?]

Focusing only on subject and predicate terms, we see that the bottom ones have tildes and the top ones don't. We need to get them into correspondence. How? Well, it occurs to me that we have an operation that allows us to add or remove tildes two at a time: contraposition. When we perform that operation, we replace the subject with the complement of the predicate (adding or removing one tilde) and we replace the predicate with the complement of the subject (adding or removing another). Now, contraposition produces equivalent sentences for A and O, but not E and I. So I can only perform it on the top sentence, "All S are P." In doing so, I produce a contrapositive that expresses the same proposition, and so must also be true. We can write it down like this:

All S are P [T]

All ~ P are ~ S [T]

No ~ S are ~ P [?]

The sentence we just wrote down still doesn't align with the target sentence at the bottom, but it's closer: they both have tildes in front of "S" and "P." Now the problem is that the "~ S" and "~ P" are in the wrong order: subject and predicate positions, respectively, in the target sentence, but the reverse in the sentence we just wrote down. We have an operation to fix that! It's called conversion: to perform it, you switch the order of subject and predicate terms. The thing is, it only works—that is, gives you an equivalent result—on E and I sentences. I can't perform conversion on the A sentence "All ~ P are ~ S" that I just wrote down at the top. But, I can perform it on the target E sentence at the bottom:

All S are P [T]

All ~ P are ~ S [T]

No ~ P are ~ S [?]

No ~ S are ~ P [?]

I did conversion, as it were, from the bottom up. Those last two E sentences are converses of one another, so they express the same proposition and will have the same truth value. If I can figure out the truth value of "No ~ P are ~ S," then I can figure out the truth value of my target sentence on the bottom; it'll be the same. And look! I'm finally in a position to do that. The two sentences in the middle, "All ~ P are ~ S" and "No ~ P are ~ S," correspond; they have the same subject and predicate. That means I can consult the Square of Opposition. I have an A sentence that's true. What about the corresponding E sentence? They're contraries, so it must be false:

All S are P [T]

All ~ P are ~ S [T]

No ~ P are ~ S [F]

No ~ S are ~ P [?]

And since the target sentence at the bottom expresses the same proposition as the one directly above it, that final question mark can also be replaced by an "F." Inference made, problem solved.

Again, this is the general pattern for making these kinds of inferences: achieve correspondence by using the three operations, then use the information encoded in the Square of Opposition.

This works most of the time, but not always. Suppose you're told that "All S are P" is true, and asked to infer the truth-value of "No P are ~ S." We can again write them out one above the other and take a look:

All S are P [T]

No P are ~ S [?]

"S" and "P" are in the wrong order, plus "S" has a tilde in front of it on the bottom but not on the top. The first thing that occurs to me to do is to get rid of that tilde. We have an operation for adding or removing one tilde at a time: obversion. I'm going to perform it on the bottom sentence. First, I change the quality: the universal negative (E) original becomes a universal affirmative (A). Then I replace the predicate with its complement: I replace "~ S" with just plain "S." This is the result:

All S are P [T]

All P are S [?]

No P are ~ S [?]

We don't have correspondence yet, but we're closer with that tilde out of the way. What next? Well, now the problem is just that "S" and "P" are in the wrong order. There's an operation for that: conversion. But—and here's the rub—we can only use conversion on E and I sentences. Now that I did obversion on the target at the bottom, the two sentences I'm left comparing are both As. I can't use conversion on an A: the result won't be equivalent.

At this point, the sensible thing to do would be to try other operations: maybe the right combination of obversion, contraposition, and possibly, eventually, on a different kind of sentence, conversion, will allow us to achieve correspondence. When making these kinds of inferences, you often have to try a variety of things before you get there. But I'm here to tell you, try what you might in this example, as many conversions, obversions, and contrapositions as you want, in whatever order: you'll never achieve correspondence. It's impossible.

So what does that mean? It means that, given the fact that "All S are P" is true, you cannot make any inference about the truth value of "No P are ~ S." The answer to the problem is: "I don't know." Remember, this kind of thing can happen; sometimes we can't make inferences about one categorical based on information about another. When we know that an I is true, for example, we can't say what the truth value of the corresponding O is; it could go either way.

That's kind of unsatisfying, though. I'm telling you that if you can't achieve correspondence—if it's impossible—that you can't make an inference. But how do you *know* that you can't achieve correspondence? Maybe, as you were laboring over the problem, you just didn't stumble on the right combination of operations in the right order. How do we know for sure that an inference can't be made?

As a matter of fact, the one step that we took in this problem puts us in a position to know just that. Compare "All S are P" with the obverse of the target sentence, "All P are S." What's the relationship between those? One is the converse of the other. We're given a true A sentence and asked to make an inference about the truth value of a sentence equivalent to its converse. But performing conversion on an A, as we established at length above, gives you a new sentence that expresses a different proposition. And this has unpredictable effects on truth value: sometimes one goes from truth to falsity, other times from truth to truth, and so on. In this case, we know that we can't know the truth value of the target sentence, because it's equivalent to the result of performing conversion on a universal affirmative, and the effects of that operation on truth value are unpredictable.

In general, you can know that the answer to one of these problems is "I don't know" if you can use the operations to get into a position where you're comparing a sentence with its converse or contrapositive when those operations don't work for the types of sentences you have. We saw this for an A and its converse. Similarly, if you have an E sentence of known truth value, and your target sentence is equivalent to its contrapositive, you know the answer is "I don't know," because contraposition performed on E sentences has unpredictable results on truth value. Same goes for I and conversion, O and contraposition.

Exercises

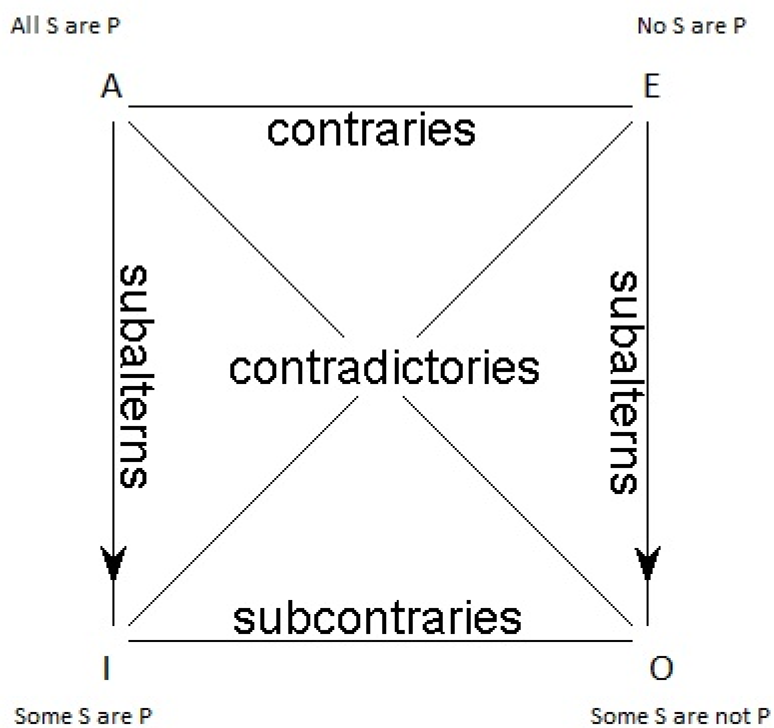
1. Suppose "All S are P" is true. Determine the truth values of the following (if possible).
   a. No S are ~ P
   b. All ~S are ~ P
   c. No ~ P are S
   d. Some ~ P are S
   e. Some ~ S are not ~ P
2. Suppose "No S are P" is true. Determine the truth values of the following (if possible).
   a. Some ~ P are not ~ S
   b. All ~ S are ~ P
   c. No ~ S are ~ P
   d. Some ~ P are S
   e. All ~ P are ~ S3.
3. Suppose "Some S are P" is true. Determine the truth values of the following (if possible).
   a. All S are ~ P
   b. Some S are not ~ P
   c. No P are S
   d. Some P are ~ S
   e. No S are ~ P4.
4. Suppose "Some S are not P" is true. Determine the truth values of the following (if possible).
   a. No S are ~ P
   b. Some S are ~ P
   c. No ~ S are P
   d. No ~ P are S

e. Some P are S

Problems with the Square of Opposition

The Square of Opposition is an extremely useful tool: it neatly summarizes, in graphical form, everything we know about the relationships among the four types of categorical proposition.

Except actually, we don't know those things. I'm sorry, but when I first presented the Square of Opposition and made the case for the various relationships it depicts, I was leading you down the proverbial primrose path. What appeared easy is in fact not as simple as it seems. Some of the relationships in the square break down under certain circumstances and force us to do some hard thinking about how to proceed. It's time to explore the "steep and thorny way" that opens before us when we dig a bit deeper into problems that can arise for the Square of Opposition.



Existential Import

To explain what these problems are, we need the concept of existential *import* (E.I. for short). E.I. is a property that propositions may or may not have. A proposition has existential import when its truth implies the existence of something. Because of what we decided to mean when we use the word "some"—namely, "there is at least one"—the particular propositions I and O clearly have E.I. For "Some sailors are not pirates" to be true, there has to *exist* at least one sailor who is not a pirate. Again, that's just a consequence of what we mean by "some."

In addition, given the relationships that are said to hold by the Square of Opposition, the universal propositions A and E also have existential import. This is because the particular propositions are subalterns. The truth of a universal proposition implies the truth of a particular one: if an A is true, then the corresponding I must be; if an E is true, then the corresponding O must be. So since the truth of universals implies the truth of particulars, and particulars have E.I., then universals imply the existence of something as well: they have existential import, too.

Problems for the Square

OK, all four of the categorical propositions have existential import. What's the big deal? Well, this fact leads to problems. Consider the proposition that all C.H.U.D.s are Republicans; also, consider the proposition that some C.H.U.D.s are not Republicans. Both of these propositions are false. That's because both of them imply the existence of things—namely, C.H.U.D.s—that don't exist. ("C.H.U.D." stands for "Cannibalistic Humanoid Underground Dweller." They're the titular scary monsters of a silly horror movie from the '80s. They're not real.) "Some C.H.U.D.s are not Republicans" claims that there exists at least one C.H.U.D. who's not a Republican, but that's not the case, since there are no C.H.U.D.s. "All C.H.U.D.s are Republicans" is also false: if it were true, its subaltern "Some C.H.U.D.s are Republicans" would have to be true, but it can't be, because it claims that there's such a thing as a C.H.U.D. (who's a Republican).

Bottom line: A and O propositions about C.H.U.D.s both turn out false. This is a problem for the Square of Opposition because A and O are supposed to be a contradictory pair; they're supposed to have opposite truth values.

It gets worse. Any time your subject class is empty—that is, like "C.H.U.D.s," it doesn't have any members—all four of the categorical propositions turn out false. This is because, as we saw, all four have existential import. But if E and I are both false, that's a problem: they're supposed to be contradictory. If I and O are both false, that's a problem: they're supposed to be subcontraries. When we talk about empty subject classes, the relationships depicted in the square cease to hold.

A Solution?

So the problems are caused by empty classes. We can fix that. We're building our own logic from the ground up here. Step one in that process is to tame natural language. The fact that natural language contains terms that don't refer to anything real seems to be one of the ways in which it is unruly, in need of being tamed. Why not simply restrict ourselves to class terms that actually refer to things, rule out empty classes? Then the square is saved.

While tempting, this solution goes too far. The fact is, we make categorical claims using empty (or at least possibly empty) class terms all the time. If we ruled these out, our ability to evaluate arguments containing such claims would be lost, and our logic would be impoverished.

One field in which logic is indispensable is mathematics. Mathematicians need precise language to prove interesting claims. But some of the most interesting claims in mathematics involve empty classes. For instance, in number theory, one can prove that there is no largest prime number—they go on forever. In other words, the term "largest prime number" refers to an empty class. If our logic ruled out empty class terms, mathematicians couldn't use it. But mathematicians are some of our best customers!

Also, physicists. Before its existence was confirmed in 2013, physicists made various claims about a fundamental particle called the Higgs boson. "Higgs bosons have zero spin," they might say, making a universal affirmative claim about these particles. But before 2013, they didn't even know if such particles existed. IF they existed, they would have zero spin (and a certain mass, etc.); the equations predicted as much. But those equations were based on assumptions that may not have been true, and so there may not have been any such particle. Nevertheless, it was completely appropriate to make claims about it, despite the fact that "Higgs boson" might have been an empty term.

We make universal claims in everyday life that don't commit us to the existence of things. Consider the possible admonition of a particularly harsh military leader: Deserters will be shot. This is a universal affirmative claim. But it doesn't commit to the existence of deserters; in fact, its very purpose is to ensure that the class remains empty!

So, empty classes have their uses, and we don't want to commit ourselves to the existence of things every time we assert a universal claim. Ruling out empty classes from our logic goes too far to save the Square of Opposition. We need an alternative solution to our problems.

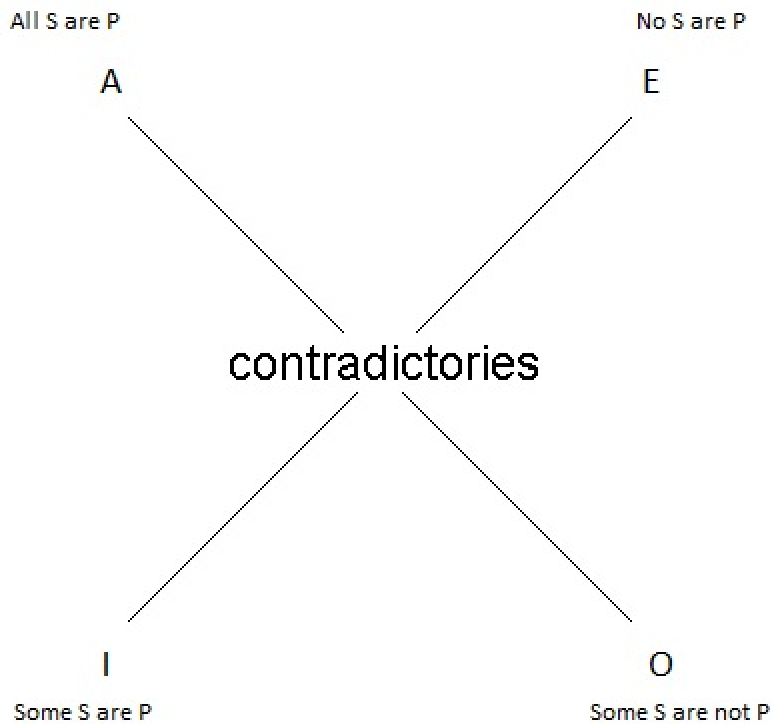
Boolean Solution

Advocated by the English logician George Boole in the 19th century, our solution to the problems raised will be to abandon the assumption that the universal propositions (A and E) have existential import, allow empty classes, and accept the consequences. Those consequences, alas, are quite dire for the traditional Square of Opposition. Many of the relationships it depicts do not hold when subject classes are empty.

First, the particular propositions (I and O) are no longer subcontraries. Since they start with the word "some," they have existential import. When their subject classes are empty, as is now allowed, they both turn out false. Subcontraries can't both be false, but I and O can both be false when we allow empty classes.

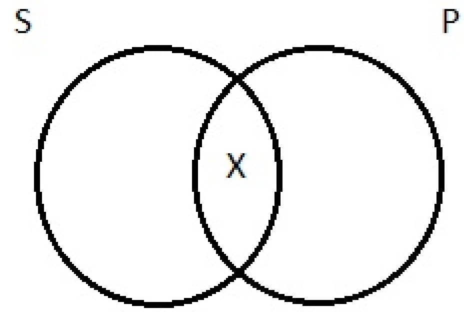
Next, the particular propositions are no longer subalterns of their corresponding universals (A and E). As we said, the universals no longer have existential import—they no longer imply the existence of anything—and so their truth cannot imply the truths of particular propositions, which do continue to have E.I. The only two relationships left on the square now are contradictoriness—between A and O, E and I—and contrariety between the two universals. And these are in conflict when we have empty subject classes. In such cases, both I and O are false, as we've said. It follows that their contradictories, A and E, must be true. But A and E are supposed to be a contrary pair; they can't both be true. So we can't keep both contrariety and contradictoriness; one must go. We will keep contradictoriness. To do otherwise would be to abandon the plain meanings of the words we're using. There's a reason I introduced this relationship first: it's the easiest to understand. If you want to contradict my universal affirmative claim that all sailors are pirates, you claim that some of them aren't; A and O are clearly contradictory. As are E and I: if you want to contradict my claim that no surfers are priests, you show me one who is. So we eliminate contrariety: it is possible, in cases where the subject class is empty, for both A and E propositions to be true.

What we're left with after making these revisions is no longer a square, but an X. All that remains is contradictoriness:

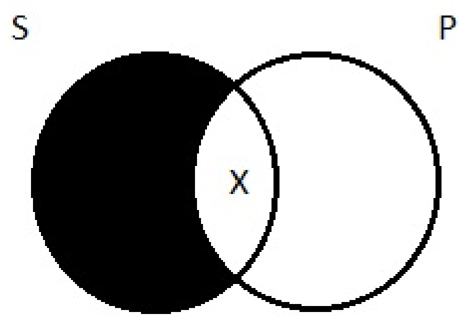


And our solution is not without awkwardness. In cases where the subject class is empty, both particular propositions (I and O) are false; their universal contradictories (E and A), then, are true in those circumstances. This is strange. Both of these sentences express truths: "All C.H.U.D.s are Republicans" and "No C.H.U.D.s are Republicans." That's a tough pill to swallow, but swallow it we must, given the considerations above. We can make it a bit easier to swallow if we say that they're true, but vacuously or trivially. That is, they're true, but not in a way that tells you anything about how things actually are in the world (the world is, after all and thankfully, C.H.U.D.-free).

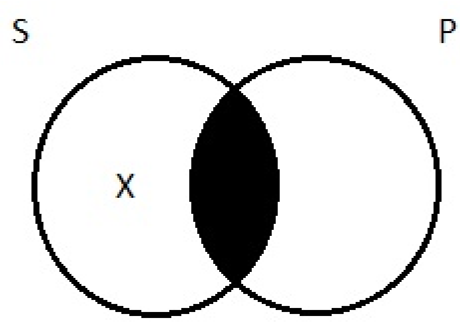
That we would end up choosing this interpretation of the categoricals, rather than the one under which universal propositions had existential import, was foreshadowed earlier, when we first introduced the four types of categorical proposition and talked about how to diagram them. We chose diagrams for A and E that did not imply the existence of anything. Recall that our way of indicating existence in Venn diagrams is to draw an X. So for a particular affirmative—some surfers are priests, say—we drew this picture (with the X being the one surfing priest we're committed to the existence of):



The diagrams for the universals (A and E), though, had no Xs in them, only shading; they don't commit us to the existence of anything. If we were going to maintain the existential import of A and E, we would've drawn different diagrams. For the universal affirmative—all logicians are jerks, say—we'd shade out the portion of the left-hand circle that doesn't overlap the right, to indicate that there's no such thing as a logician who's not a jerk. But we would also put an X in the middle region, to indicate that there is at least one logician who is (existential import):



And for the universal negative—no women are priests, say—we would shade out the middle region, to indicate that there's nothing that's both a woman and a priest. But we would also put an X in the left-hand circle, to indicate that there's at least one woman who's not a priest:



This interpretation of the universal propositions, according to which they have existential import, is often called the "Aristotelian" interpretation (as opposed to our "Boolean" interpretation, according to which they do not).[9] Which interpretation one adopts makes a difference. There are some arguments that the two

9. It is not clear, however, that it is correct to attribute this view to Aristotle. While he clearly did believe that universal affirmative (A) propositions had

interpretations evaluate differently: on the Aristotelian view, they are valid, but on the Boolean view, they are not. We will stick to the Boolean interpretation of the universals, according to which they do not have existential import.

His rendering of the particular negative (O) was "Not all S are P," which could be (trivially, vacuously) true when S is empty. In that case, O's being the subaltern of E does not force us to attribute Existential Import to the latter.

Categorical Syllogisms

As we've said, Aristotelian logic limits itself to evaluating arguments all of whose propositions—premises and conclusion—are categorical. There is a further restriction: Aristotelian logic only evaluates categorical syllogisms. These are a special kind of argument, meeting the following conditions:

> A *categorical syllogism* is a deductive argument consisting of three categorical propositions (two premises and a conclusion); collectively, these three propositions feature exactly three classes; each of the three classes occurs in exactly two of the propositions.

That's a mouthful, but an example will make it clear. Here is a (silly) categorical syllogism:

1. All chipmunks are Republicans.
2. Some Republicans are golfers.

So, some chipmunks are golfers.

This argument meets the conditions in the definition: it has three propositions, there are exactly three classes involved (chipmunks, Republicans, and golfers), and each of the three classes occurs in exactly two of the propositions (check it and see).

There is some special terminology for the class terms and premises in categorical syllogisms. Each of the three class terms has a special designation. The so-called *major term* is the term that appears in predicate position in the conclusion; in our silly example, that's "golfers." The *minor term* is the term that appears in subject

---

existential import, it's not clear that he thought the same about universal negatives. For discussion, see Parsons, Terence, "The Traditional Square of Opposition," The Stanford Encyclopedia of Philosophy (Summer 2015 Edition), Edward N. Zalta (ed.), URL = http://plato.stanford.edu/archives/sum2015/entries/square/

position in the conclusion; in our example, that's "chipmunks." The *middle term* is the other one, the one that appears in each of the premises; in our example, it's "Republicans."

The premises have special designations as well. The *major premise* is the one that has the major term in it; in our example, that's "Some Republicans are golfers." The *minor premise* is the other one, the one featuring the minor term; in our example, it's "All chipmunks are Republicans."

Final restriction: categorical syllogisms must be written in standard form. This means listing the premises in the correct order, with the major premise first and the minor premise second. If you look at our silly example, you'll note that it's not in standard form. To fix it, we need to reverse the order of the premises:

1. Some Republicans are golfers.
2. All chipmunks are Republicans.

So some chipmunks are golfers.

An old concern may arise again at this point: in restricting itself to such a limited class of arguments, doesn't Aristotelian logic run the risk of not being able to evaluate lots of real-life arguments that we care about? The response to this concern remains the same: while most (almost all) real-life arguments are not presented as standard form categorical syllogisms, a surprising number of them can be translated into that form. Arguments with more than two premises, for example, can be rewritten as chains of two-premise sub-arguments. As was the case when we raised this concern earlier, we will set aside the messy details of exactly how this is accomplished in particular cases.

Logical Form

As we said at the outset of our exploration of deductive logic, there are three things such a logic must do: (1) tame natural language, (2) precisely define logical form, and (3) develop a way to test logical forms for validity. Until now, we've been concerned with the first step. It's (finally) time to proceed to the second and third.

> The logical form of a categorical syllogism is determined by two features of the argument: its *mood* and its *figure*.

First, mood. The mood of a syllogism is determined by the types of categorical propositions contained in the argument, and the order in which they occur. To determine the mood, put the argument into standard form, and then simply list the types of categoricals (A, E, I, O) featured in the order they occur. Let's do this with our silly example:

1. Some Republicans are golfers.
2. All chipmunks are Republicans.

So some chipmunks are golfers.

From top to bottom, we have an I, an A, and an I. So the mood of our argument is IAI. It's that easy. It turns out that there are 64 possible moods—64 ways of combining A, E, I, and O into unique three-letter combinations, from AAA to OOO and everything in between.

The other aspect of logical form is the argument's figure. The figure of a categorical syllogism is determined by the arrangement of its terms. Given the restrictions of our definition, there are four different possibilities for standard form syllogisms. We will list them schematically, using these conventions: let "S" stand for the minor term, "P" stand for the major term, and "M" stand for the middle term. Here are the four figures:

(i)	MP	(ii)	PM	(iii)	MP	(iv)	PM
	SM		SM		MS		MS
	SP		SP		SP		SP

Again, the only thing that determines figure is the arrangement of terms—whether they appear in subject or predicate position in their premises. In our schemata, that the letter is listed first indicates that the term appears in subject position; that it appears second indicates that it's in predicate position. So, in the first figure, in the major premise (the first one), the middle term (M) is in subject position and the major term (P) is in predicate position. Notice that for all four figures, the subject and predicate of the conclusion remains the same: this is because, by definition, the minor term (S) is the subject of the conclusion and the major term (P) its predicate.

Returning to our silly example, we can determine its figure:

1. Some Republicans are golfers.
2. All chipmunks are Republicans.

So, some chipmunks are golfers.

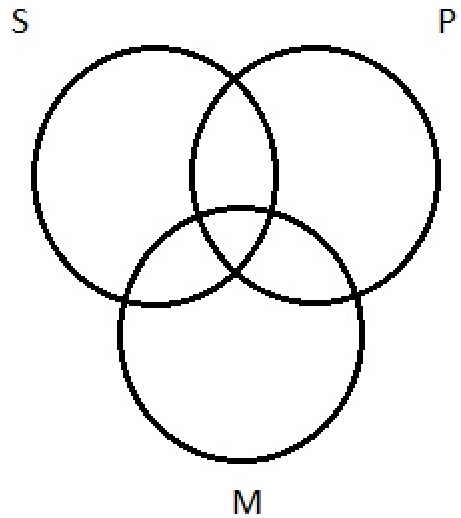
Perhaps the easiest thing to do is focus on the middle term, the one that appears in each of the premises—in this case, "Republicans." It occurs in subject position in the major premise, then predicate position in the minor premise. Scanning the four figures, I just look for the one that has "M" listed in first position on the top, then second position in the middle. That's the first figure. So the mood of our sample argument is IAI, and it's in the first figure. Logical form is just the mood and figure, and conventionally, we list logical forms like this: IAI-1 (the mood, a dash, then a number between 1 and 4 for the figure).

There are 4 figures and 64 moods. That gives us 256 possible logical forms. It turns out that only 15 of these are valid. We need a way to test them. It is to that task we now turn.

The Venn Diagram Test for Validity

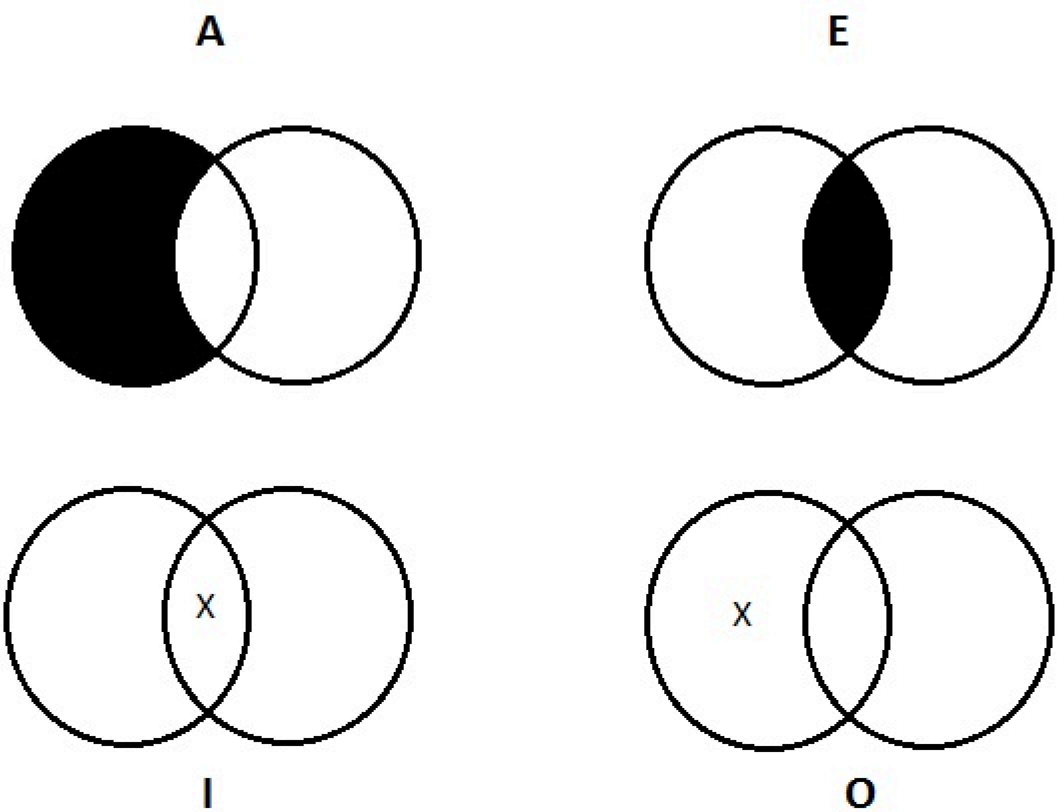
To test syllogistic forms for validity, we proceed in three steps:

Step 1. Draw three overlapping circles, like this:



That gives us one circle for each of the three terms in the syllogism: minor (S), major (P), and middle (M).

Step 2. Depict the assertions made by the premises of the syllogism on this diagram, using shading and Xs as appropriate, depicting the individual A, E, I, or O propositions in the usual way:



Each of the premises will be a proposition concerning only two of the three classes (S, P, and M). The major premise will concern M and P (in some order); the minor premise will concern M and S (in some order). How the circles will be labeled (with S, M, P) will depend on these particulars.

Step 3. After the premises have been depicted on the three-circle diagram, we look at the finished product and ask, "Does this picture entail the truth of the conclusion?" If it does, the form is valid; if it does not, it is invalid.

In the course of running the test, we will keep two things in mind—one rule of thumb and one convention:

Rule of Thumb: In step 2, depict universal (A and E) premises before particular (I and O) ones (if there's a choice).

Convention: In cases of indeterminacy, draw Xs straddling boundary lines.

We need to explain what "indeterminacy" amounts to; we will in a moment. For now, to make all this more clear, we should run through some examples.

Let's start at the beginning (alphanumerically): AAA-1. We want to test this syllogistic form for validity. What does an argument of this form look like, schematically? Well, all three of its propositions are universal affirmatives, so they're all of the form All __ are __. We have:

1. All __ are __
2. All __ are __

So, all __ are __

That's what the mood (AAA) tells us. We have to figure out how to fill in the blanks with S, P, and M. The figure tells us how to do that. AAA-1: so, first figure. That looks like this:

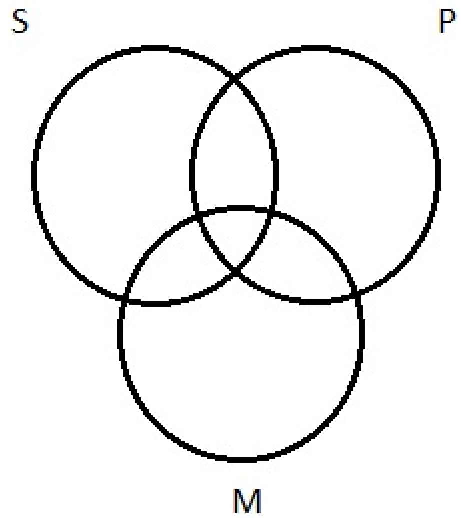
(i) MP
SM
———
SP

So AAA-1 can be schematically rendered thus:

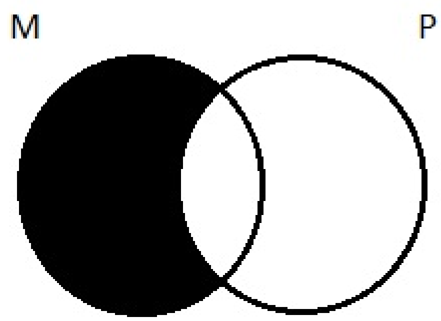
1. All M are P.
2. All S are M.

So, all S are P.

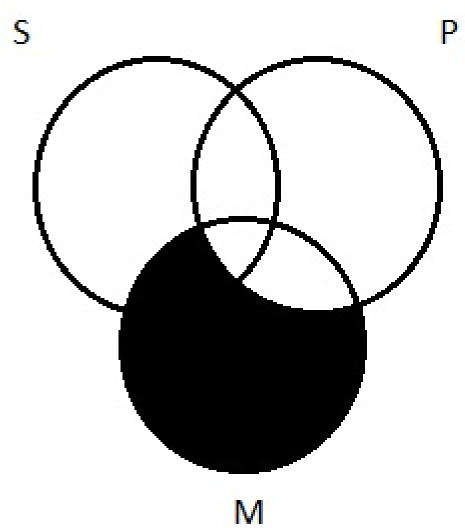
To test this form for validity, we start with step 1, and draw three circles:



In step 2, we depict the premises on this diagram. (We're supposed to keep in mind the rule of thumb that, given a choice, we should depict universal premises before particular ones, but since both of the premises are universals, this rule does not apply to this case.) We can start with the major premise: All M are P. On a regular two-circle Venn diagram, that would look like this:

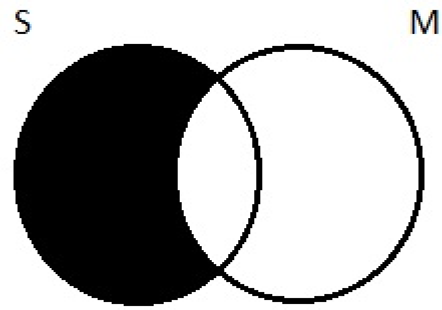


The trick is to transfer this two-circle diagram onto the three-circle one. In doing so, we keep in mind that all the parts of M that are outside of P must be shaded. That gives us this:

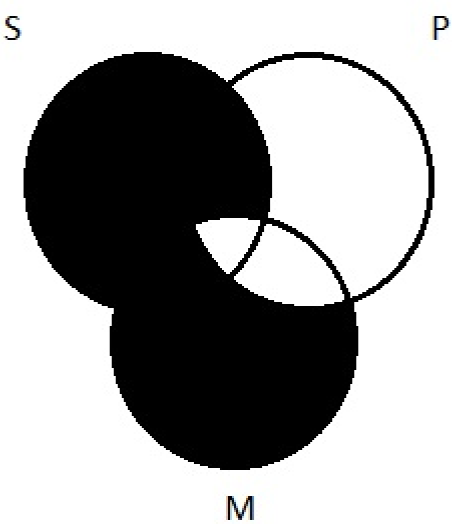


Note that in the course of shading out the necessary regions of M, we shaded out part of S. That's OK. Those members of the S class are Ms that aren't Ps; there's no such thing, so they have to go.

Next, we depict the minor premise: All S are M. With two circles, that would look like this:

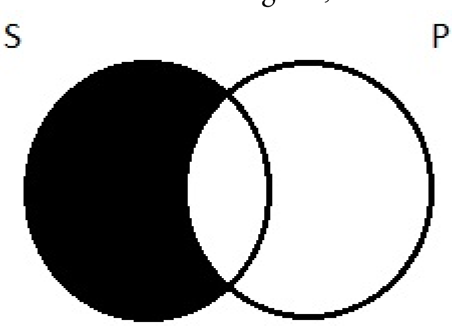


Transferring that onto the three-circle diagram means shading all the parts of S outside of M:



Step 2 is complete! We have depicted the assertions made by the premises.

In step 3 we ask whether this diagram guarantees the truth of the conclusion. Well, our conclusion is "All S are P." In a two-circle diagram, that looks like this:



Does our three-circle diagram guarantee the truth of "All S are P"? Focusing on the S and P circles, and comparing the two diagrams, there's a bit of a difference: part of the area of overlap between S and P is shaded out in our three-circle diagram, but it isn't in the two-circle depiction. But that doesn't affect our judgment about whether the diagram guarantees "All S are P." Remember, this can be thought of as a claim that a certain kind of thing doesn't exist—an S that's outside the P circle. If there are any Ss (and there may not be), they will also be Ps. Our three-circle diagram does in fact guarantee this. There can't be an S that's not a P; those areas are shaded out. Any S you find will also be a P; it'll be in that little region in the center where all three circles overlap.

So, since the answer to our question is "yes," the syllogistic form AAA-1 is valid.[10]

We should reflect for a moment on why this method works. We draw a picture that depicts the assertions made by the premises of the argument. Then we ask whether that picture guarantees the conclusion. This should sound familiar. We're testing for validity, and by definition, an argument is valid if and only if its premises guarantee its conclusion; that is, IF the premises are true, then the conclusion must also be true. Our method mirrors the definition. When we depict the premises on the three-circle diagram, we're drawing a picture of what it looks like for the premises to be true. Then we ask, about this picture—which shows a world in which the premises are true—whether it forces us to accept the conclusion—whether it depicts a world in which the conclusion must be true. If it does, the argument is valid; if it doesn't, then it isn't. The method follows directly from the definition of validity.

To further illustrate the method, we should do some more examples. AII-3 is a useful one. The mood tells us it's going to look like this:

1. All __ are __
2. Some __ are __

So, some __ are __

And we're in the third figure:

10. Trivial fact: all the valid syllogistic forms were given mnemonic nicknames in the Middle Ages to help students remember them. AAA-1 is called "Barbara." No, really. All the letters in the name had some meaning: the vowels indicate the mood (AAA); the other letters stand for features of the form that go beyond our brief investigation into Aristotelian logic.

(iii) MP
MS
SP

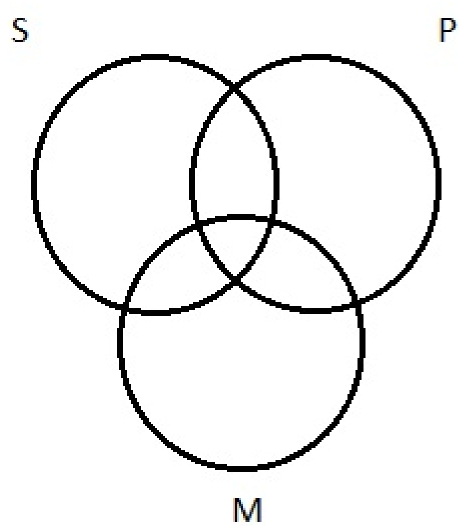
So we fill in the blanks to get the schematic form:

1. All M are P
2. Some M are S

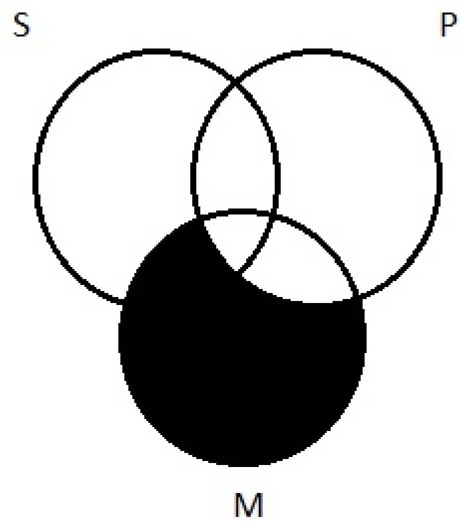
So, Some S are P

Now we can test this form.

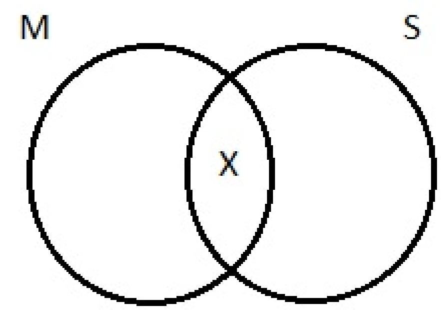
Step 1: We start the test of this form with the blank three-circle diagram:



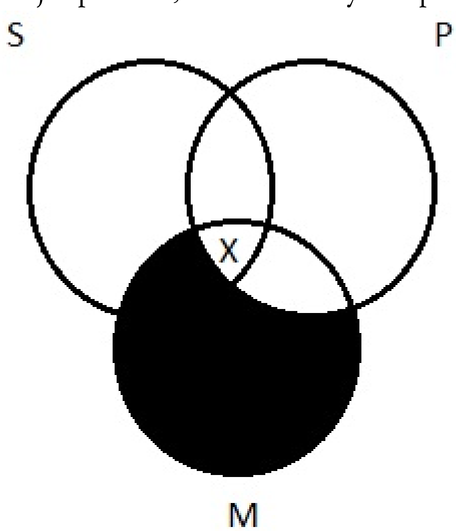
Step 2: Depict the premises. And here, our rule of thumb applies: depict universals before particulars. The major premise is a universal (A) proposition; the minor premise is a particular (I). So we depict the major premise first. That's "All M are P." We did this already. Recall that Barbara has the same major premise. So depicting that on the diagram gives us this:



Next, the minor premise: Some M are S. Recall, with particular propositions, we depict them using an X to indicate the thing said to exist. This proposition asserts that there is at least one thing that is both M and S:



We need to transfer this to the three-circle diagram. We need an X that is in both the M and S circles. If we look at the area of overlap between the two, we see that part of it has been shaded out as the result of depicting the major premise, so there's only one place for the X to go:

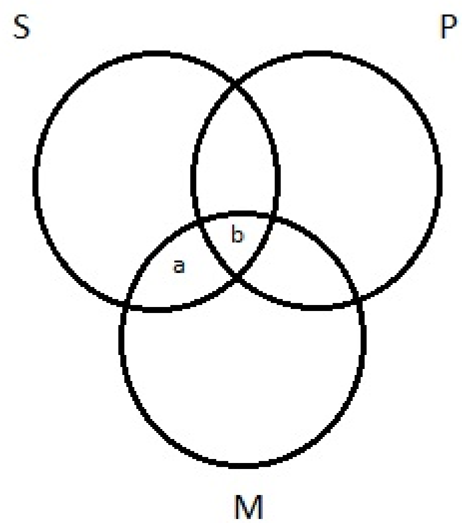


Step 2 is complete! The premises are depicted.

So we proceed to step 3 and ask, "Does this picture guarantee the conclusion?" The conclusion is "Some

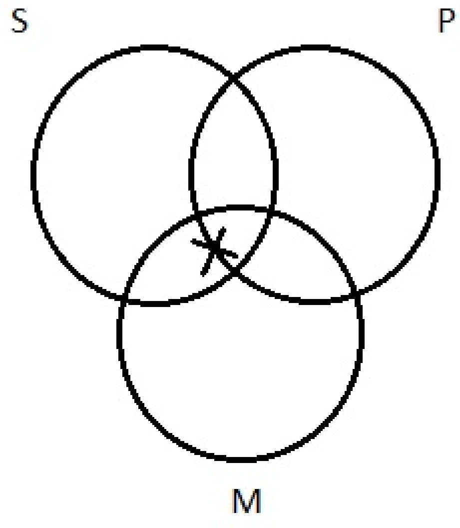
S are P"; that's an assertion that there is at least one thing that is both S and P. Is there? Yes! That X that we drew in the course of depicting the minor premise is in the sweet spot—the area of overlap between S and P. It guarantees the conclusion. The argument is valid.[11]

That's another successful use of the Venn diagram test for validity, but I want to go back and revisit some of it. *I want us to reflect on why we have the rule of thumb to depict universal premises before particular ones.* Remember, we had the universal major premise "All M are P" and the particular minor premise "Some M are S." The rule of thumb had us depict them in that order. Why? What would have happened had we done things the other way around? We would have started with a blank three-circle diagram and had to depict "Some M are S" on it. That means an X in the area of overlap between M and S. That area, though, is divided into two sub-regions (labeled "a" and "b"):



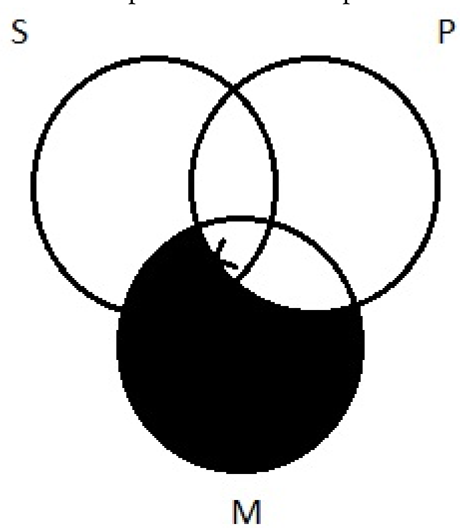
Where do I put my X—in region a or b? Notice, it makes a difference: if I put the X is region a, then it's outside the P circle; if I put it in region b, then it's inside the P circle. The question is: "Is this thing that the minor premise says exists a P or not a P?" I'm depicting a premise that only asserts "Some M are S." That premise says nothing about P. It's silent on our question; it gives us no guidance about how to choose between regions a and b. What to do? This is one of the cases of "indeterminacy" that we mentioned earlier when we introduced a convention to keep in mind when running the test for validity: In cases of indeterminacy, draw Xs straddling boundary lines. We don't have any way of choosing between regions a and b, so when we draw our X, we split the difference:

11. If you're curious, its mnemonic nickname is "Datisi." Weird, I know; it was the Middle Ages.

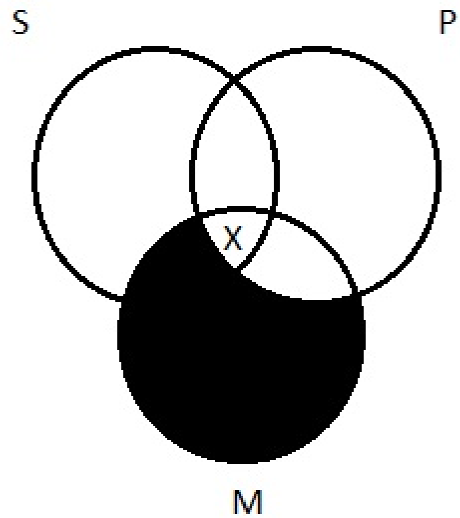


This drawing indicates that there's an X in there somewhere, either inside or outside the P circle; we don't know which.

And now we see the reason for our rule of thumb—depict universals before particulars. Because if we proceed to depict the universal premise "All M are P," we shade thus:



The shading erased half our X. That is, it resolved our question of whether the X should go in the P circle: it should. So now we have to go back and erase the half-an-X that's left and re-draw the X in that center region in order to end up with the finished diagram we arrived at earlier:



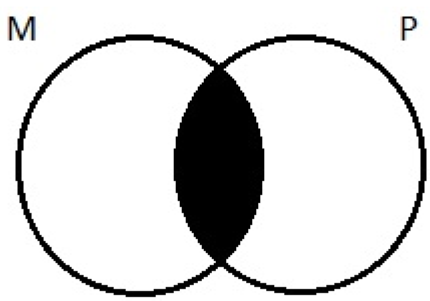
We would've saved ourselves the trouble had we just followed the rule of thumb to begin with and depicted the universal before the particular—shading before the X. That's the utility of the rule: sometimes it removes indeterminacy that would otherwise be present.

One more example to illustrate how this method works. Let's test EOI-1. Noting that in the first figure the middle term is first subject and then predicate, we can quickly fill in the schema:

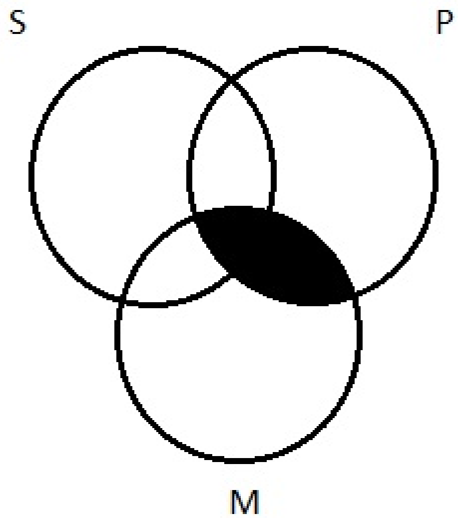
1. No M are P
2. Some S are not M.

So, some S are P.

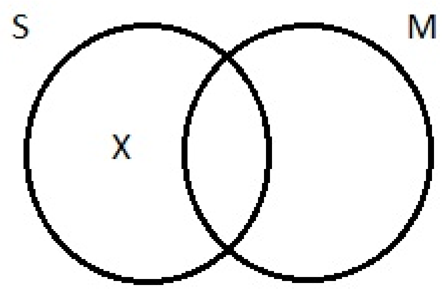
Following the rule of thumb, we depict the universal (E) premise first. No M are P asserts that there is nothing that is in both of those classes. The area of overlap between them is empty. With two circles, we have this:



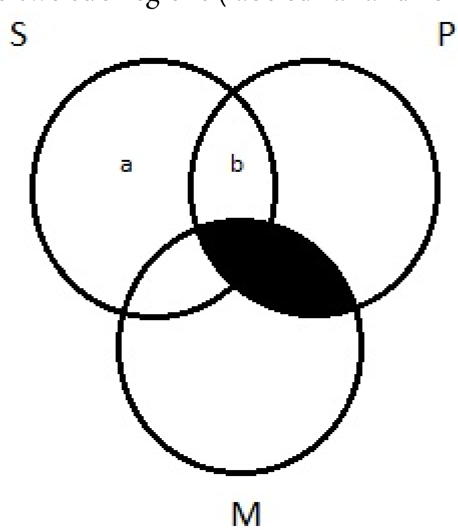
Transferring this onto the three-circle diagram, we shade out all the area of overlap between the M and P circles (clipping off part of S along the way):



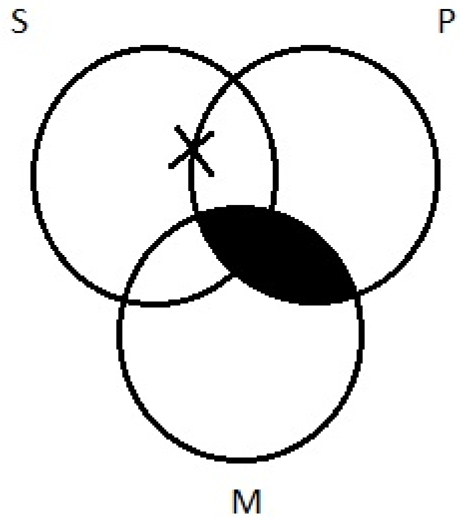
Next, the particular (O) premise: Some S are not M. This asserts the existence of something—namely, a thing that is an S but not an M. We need an X in the S circle that is outside the M circle:



Moving to the three-circle diagram, though, things get messy. The area of S that's outside of M is divided into two sub-regions (labeled "a" and "b"):



We need an X somewhere in there, but do we put it in region a or region b? It makes a difference: if we put it in region b, it is a P; if we put it in region a, it is not. This is the same problem we faced before. We're depicting a premise—Some S are not M—that is silent on the question of whether or not the thing is a P. Indeterminacy. We can't decide between a and b, so we split the difference:



That X may be inside of P, or it may not; we don't know. This is a case in which we followed the rule of thumb, depicting the universal premise before the particular one, but it didn't have the benefit that it had when we tested AII-3: it didn't remove indeterminacy. That can happen. The rule of thumb is in place because it sometimes removes indeterminacy; it doesn't always work, though.

So now that we've depicted the premises, we ask whether they guarantee the conclusion. Is the world depicted in our diagram one in which the conclusion must be true? The conclusion is "Some S are P": it asserts that there is at least one thing that is both S and P. Does our picture have such a thing? There's an X in the picture. Does it fit the bill? Is it both S and P? *Well, uh…maybe?* That X may be inside of the area of overlap between S and P; then again, it may not be.

Oy. What do we say? It's tempting to say this: we don't know whether the argument is valid or not; it depends on where that X really is. But that's not the correct response. Remember, we're testing for validity—for whether or not the premises *guarantee* the conclusion. We can answer that question: they don't. For a guarantee, we would need an X in our picture that is *definitely* inside that middle region. We don't have such an X. These premises leave open the possibility that the conclusion is true; they don't rule it out. But that's not enough for validity. For an argument to be valid, the premises must *necessitate* the conclusion, force it on us. These do not. Thus, the form EOI-1 is not valid.

Exercises

1. Identify the logical form of the following arguments.
    a. Because some Wisconsinites are criminals and all criminals are scoundrels, it follows that some scoundrels are Wisconsinites.
    b. No surfers are priests, because all priests are men and some surfers are not men.
    c. Some authors are feminists, since some women are authors and some women are feminists.
    d. All mosquitoes are potential carriers of disease; therefore some mosquitoes are a menace to society, since all potential carriers of disease are a menace to society.

   e. Because some neo-Nazis are bloggers, some neo-Nazis are not geniuses, since no geniuses are bloggers.

2. Test the following syllogistic forms for validity.
   a. EAE-2
   b. EAE-3
   c. OAO-3
   d. EIO-4
   e. AOO-4
   f. IAI-1
   g. AII-1
3. Test the following arguments for validity.
   a. Some pirates are mercenaries; hence, some sailors are pirates, because all sailors are mercenaries.
   b. Some women are not nuns, but all nuns are sweethearts; it follows that some women are not sweethearts.
   c. Some Republicans are not politicians, for some Republicans are not Christians, and some Christians are not politicians.
4. Test the arguments in Exercise 1 for validity.

5.

SYMBOLIC TRANSLATIONS

Why Translate Statements?

We have seen that many arguments can appear compelling even when they lack logical force. The strategies employed to dupe us (or simply those that catch us off guard) are embedded in what we will call "natural languages." These arguments are expressed in existing spoken, written, or signed languages. This text is written in English, so we have thus far looked at arguments in that natural language. And of course, we see that it can get us in an awful lot of trouble.

The difficulty we often find is that natural language use is highly complex. As J. L. Austin pointed out long ago, *we do things with words*. While we want to be sensitive to the nefarious things people do with language (see: Informal Fallacies), we also want to focus our efforts on one of the most important things we do with language: understand the meaning of claims.

You might think that this is obvious: "*I mean, I speak* English*, so I understand it when I hear it*." How we wish that were so…

I often think studying logic might be easier if it were conveyed in an ancient Germanic tongue or some rare dialect of the Inuit people. If it were, most of us would probably pay very close attention to what was being said. *We would try*. Sadly, because most arguments are conveyed in a language we speak, we tend to think we do not need to try—the job of understanding has already been done! This is not so. Yet there is a clue here to how we can help ourselves avoid this error. If someone *did* convey an argument in the Inuit language, *we would need a translation*—something that helps us understand what was originally said.

Welcome to symbolic translations! We translate statements in one language (English) to another (our symbolic language) for the same reason we translate the Inuit arguments: to understand them. This only seems odd at first because of our false assumption that we already understand what people are saying when they are speaking a language we speak. Many of us know this to be false because we have seen people argue from a third-party perspective. We were not in it, we did not have a dog in that fight, and from that perspective we often note how people frequently miss the whole point of what the other person was saying. Two people, apparently speaking the same language, can fail to communicate effectively with one another…*and the same thing is happening in our own heads*.

Once we realize that there is a part of your brain that is trying to communicate with another part of your brain, we should recognize the danger of misunderstanding. What one part says, the other part *only believes* it understands. A good translation might help the other part grasp things more clearly. A good translator is needed, one well-trained at identifying what is needed in the moment and to the greatest benefit of the recipient. *That will be you*.

Starting with Statements

We begin the study of logic by building a precise logical language. This will allow us to do at least two things: first, to say some things more precisely than we otherwise would be able to do; second, to study reasoning. We will use a natural language—English—as our guide, but our logical language will be *far simpler*, far weaker, *but more rigorous* than English.

We must decide where to start. We could pick just about any part of English to try to emulate: names, adjectives, prepositions, general nouns, and so on. But it is traditional, and as we will see, quite handy, to begin with whole sentences. For this reason, the first language we will develop is called "the propositional logic." It is also sometimes called "the sentential logic" or even "the sentential calculus." These all mean the same thing: the logic of sentences.

In this propositional logic, the smallest independent parts of the language are sentences. Throughout this book, I will assume that sentences and propositions are the same thing in our logic, and I will use the terms "sentence" and "proposition" interchangeably. Our preferred term will be "statements" to mean the same thing. The term "statements" is more precise and, as we will shortly see, helps clarify distinctions between some natural language "sentences" and proper propositions of logic.

There are of course many kinds of sentences. To take examples from our natural language, these include:

What time is it?

Open the window.

Damn you!

I promise to pay you back.

It rained in Central Park on June 26, 2015.

We could multiply such examples. Sentences in English can be used to ask questions, give commands, curse or insult, form contracts, and express emotions. But the last example above is of special interest because it aims to describe the world. Such sentences, which are sometimes called "declarative sentences," or better yet, "declarative statements," will be our model sentences for our logical language. We know a declarative sentence when we encounter it because it can be either true or false. This is why we will benefit from the term "statement" to distinguish these important uses of language from other uses that have a different relationship to truth. Consider some of the above examples of English sentences:

What time is it?

Damn you!

If I ask you if either of these are true or false, you might look at me strangely. They are, of course, English sentences, but they bear a different relationship to truth. The first does not bear any truth or falsehood—rather, it inquires into the truth. The second also does not bear any truth, rather (at best, in its most classic sense), it seeks to change the truth of a future state. As we move forward, we will do well to keep in mind that logicians are concerned only with those sentences that can be true or can be false. So when we say "sentence" we mean "statement."

Truth Value and Precision in Sentences

We want our logic of declarative sentences to be precise. But what does this mean? We can help clarify how we might pursue this by looking at sentences in a natural language that are perplexing, apparently because they are not precise. Here are two.

Tom is kind of tall.

When Karen had a baby, her mother gave her a pen.

We have already observed that an important feature of our declarative sentences is that they can be true or false. We call this the **"truth value"** of the sentence. These two sentences are perplexing because their truth values are unclear. The first sentence is vague; it is not clear under what conditions it would be true, and under what conditions it would be false. If Tom is six feet tall, is he kind of tall? There is no clear answer. The second sentence is ambiguous. If "pen" means writing implement, and Karen's mother bought a playpen for the baby, then the sentence is false. But until we know what "pen" means in this sentence, we cannot tell if the sentence is true.

Since ancient times, philosophers have believed that we will deceive ourselves, and come to believe untruths, if we do not accept a principle sometimes called "bivalence," or a related principle called "the principle of non-contradiction." Bivalence is the view that there are only two truth values (true and false) and that they exclude each other. The principle of non-contradiction states that you have made a mistake if you both assert and deny a claim.

We can take these observations for our guide: we want our language to have no vagueness and no ambiguity. In our propositional logic, this means we want it to be the case that each sentence is either true or false. It will not be kind of true, or partially true, or true from one perspective and not true from another.

We can formulate our own revised version of the principle of bivalence, which states that:

Principle of Bivalence: Each sentence of our language must be either true or false: not both, not neither.

Some readers may be thinking: what if I reject bivalence, or the principle of non-contradiction? If you have doubts about bivalence, or the principle of non-contradiction, stick with logic. That is because we could develop a logic in which there were more than two truth values. Logics have been created and studied in which we allow for three truth values, or continuous truth values, or stranger possibilities. The issue for us is that we must start somewhere, and the principle of bivalence is an intuitive way and—it would seem—the simplest way to start with respect to truth values. Learn basic logic first, and then you can explore these alternatives.

The Correspondence Theory of Truth

Our starting point will embrace both principles of bivalence and non-contradiction. We will also embrace an intuitive concept of truth value[1]: correspondence with the facts of the world. This is easiest to understand if we think of our terms as technical terms. The term "true" is not the same as the term "factual." One (truth value) is a property of a statement. The other (fact) is a property of the world. Put differently:

> *The world be whatever it may be. What we say in an attempt to declare that things are so is an entirely different matter.*

So, we should recognize that when we make a statement, we are not creating the world. We are attempting to capture it, to describe it, to understand it. Our words, our statements, are accountable to the world. They try to describe the world in a certain way, and they either get it right or they muck it up in some way. The world be what it be(those are the facts)—our attempts to describe it are the things that can succeed or fail (those are things that we call true or false).

This correspondence theory of truth is pretty intuitive and common. Most of us walk around on a daily basis adhering to this notion of truth value.

Ken: *Honey, where are my keys? I left them on the kitchen counter last night*.

Molly: *No dear, they're not there. They're in your coat pocket*.

This might be how Ken sees things, and this is his attempt to describe the world. Did he get it right? You only have to go look at the relevant part of the world to find out. Ken's statement is "false" because the world is not as he describes it. Molly's statement is "true" because the world is as she describes it. The world itself is *neither* true nor false—our statements have that property. The world just has facts; it is in a certain way, in a particular state, and we try to capture those facts with our language. So we can say again: **only statements have the property of truth value**.

We all have some grasp of what "true" means, and this grasp will be sufficient for our development of propositional logic.

Analyzing Statement Structure

Our language will be concerned with declarative statements: statements that are either true or false, never both, and never neither. Here are some example statements:

2+2=4.

Malcolm Little is tall.

If Lincoln wins the election, then Lincoln will be president.

---

1. We have already seen this concept of truth value in Chapter 1, but this is a good time to revisit it.

The Earth is not the center of the universe.

These are all declarative sentences. These all appear to satisfy our principle of bivalence. But they differ in important ways. The first two sentences do not have sentences as parts. For example, try to break up the first sentence. "2+2" is a function. "4" is a name. "=4" is a meaningless fragment, as is "2+." Only the whole expression, "2+2=4," is a sentence with a truth value. The second sentence is similar in this regard. "Malcolm Little" is a name. The "is tall" is an adjective phrase (we will discover later that logicians call this a predicate"). "Malcolm Little is" or "is tall" are fragments; they have no truth value. Only "Malcolm Little is tall" is a complete sentence.

Atomic Statements

The first two example sentences above are of a kind we call "atomic statements." The word "atom" comes from the ancient Greek word "atomos," meaning "cannot be cut." When the ancient Greeks reasoned about matter, for example, some of them believed that if you took some substance, say a rock, and cut it into pieces, then cut the pieces into pieces, and so on, eventually you would get to something that could not be cut. In logic, our atomic statements cannot be cut up into smaller parts without losing something of great significance: truth value. They are the smallest possible thing that retains the property of truth value.

In logic, the idea of an atomic sentence is of a sentence that can have no parts *that are* sentences. We should see, much like their physical namesake, that atomic statements clearly do have parts—but those parts are not in themselves statements (they are "subatomic" particles that cannot directly bear truth value). Noticing these parts can be a helpful way to identify an atomic statement, which always has the form:

One Subject + One Predicate

Consider the following:

Kevin likes weightlifting.

Carl enjoys taking long walks along the River Thames.

The newest Shelby Mustang has over 600 horsepower.

My cousin Mary won a baking contest last year.

Each of these has one subject and one thing said of that subject. Note, however, that we are not saying "one word" is the subject of the statement. Subjects are often expressed in a single "term" that requires several words to detail. "My cousin Mary" is a single term, as is "the newest Shelby Mustang." The speaker is only referring to *one thing*, even though they use several words to do it.

If we have a statement with more than one subject or more than one predicate, then we have something larger than an atomic statement. We'll get to those in a bit. For now, we can see that statements like the following have more than one subject or more than one predicate:

Tom is a cat, and he eats mice.

Missy and Andrew got As on the exam.

My house is small but cozy.

The first and third have only one subject (one “thing” about which the statement attempts to describe). However, they also have two predicates (two different descriptions of that same thing). In the second example, there is only one description (one predicate) given, but it is given to two distinct subjects.

To express our understanding of these larger statements in our symbolic language will require a bit of work. However, to express our understanding of an atomic statement like “Tom is a cat” is easy. After all, the smaller bits in this sort of statement do not have a truth value, so for now they can be suppressed into one “atom” of a meaningful statement.

It is traditional in logic to use upper case letters from P on (P, Q, R, S….) to stand for atomic sentences. Thus, instead of writing

“Malcolm Little is tall.”

We could write

P

If we want to know how to translate P to English, we can provide a translation key. Similarly, instead of writing

“Malcolm Little is a great orator.”

We could write

Q

And so on. Of course, written in this way, all we can see about such a sentence is that it is a sentence, and that perhaps P and Q are different sentences. But for now, these will be sufficient.

Note that not all sentences are atomic. Recall the example given previously: *If Lincoln wins the election, then Lincoln will be president.*

This contains parts that *are* sentences. It contains the atomic sentence “Lincoln wins the election” and the atomic sentence “Lincoln will be president.” We could represent this whole sentence with a single letter. That is, we could let

“If Lincoln wins the election, Lincoln will be president.”

be represented in our logical language by

S

However, this would have the disadvantage that it would hide some of the sentences that are inside this sentence, and it would hide their relationship. We would have a translation that suppresses important meaning that is critical to a logical evaluation of the original statement. Our symbolic language would tell us more if we could capture the relationship between the parts of this sentence, instead of hiding it. *So this would be a **very poor** translation of this statement.*

Syntax and Semantics

An important and useful principle for understanding a language is the difference between syntax and semantics. "Syntax" refers to the "shape" of an expression in our language. It does not concern itself with what the elements of the language mean, just specifies how they can be written out.

We can make a similar distinction (though not exactly the same) in a natural language. This expression in English has an uncertain meaning, but it has the right "shape" to be a sentence:

Colorless green ideas sleep furiously.

In other words, in English, this sentence is syntactically correct, although it may express some kind of meaning error.

An expression made with the parts of our language must have correct syntax in order to be a sentence. Sometimes, we also call an expression with the right syntactic form a "well-formed formula" or WFF for short.

We contrast syntax with semantics. "Semantics" refers to the meaning of an expression of our language. Semantics depends upon the relationship of that element of the language to something else. For example, the truth value of the sentence "The Earth has one moon" depends not upon the English language, but upon something exterior to the language (i.e., the facts of the world that be this way). Since the self-standing elements of our propositional logic are sentences, and the most important property of these is their truth value, the only semantic feature of sentences that will concern us in our propositional logic is their truth value.

We should note that there are other semantic features of sentences that we will ignore in our efforts to translate into a symbolic language. This is an inevitable consequence of going from one language to another. Anyone who has ever translated from one language to another has come across times when they just shrug and say, "*Sorry, there really isn't a way to convey the same meaning of what they said in your language; there just isn't a word for it in your language*." In these instances, we say some of the meaning was "lost in translation." In logic, we sometimes refer to this as *the sense* of an expression. As we will see, this "sense" is frequently lost, yet the truth value remains.

We have so far introduced atomic sentences. The syntax for an atomic sentence is trivial. If P is an atomic sentence, then it is syntactically correct to write down

P

By saying that this is syntactically correct, we are not saying that P is true. Rather, we are saying that P is a sentence. Fortunately, in a proper translation we will always provide a key to understand what each atomic statement means in English. These keys have different names in different logic texts. We see them referred to as:

Translation Scheme

Scheme of Abbreviation

Universe of Discourse

These terms all mean the same thing: a key to tell us what each atomic (each capital letter) means in the present translation. If you are provided a translation scheme, you should use it. However, if you are given no translation scheme, *you should most definitely write it out completely* so that it accompanies your translation of the larger statements.

Compound Statements

Statements larger than atomics are usually composed of multiple atomic statements. These go by different names. Generally, we refer to these as compound statements, but they are also called "complex" statements and "molecular" statements (because they are composed of more than one atomic, a play on the physical analogy). However, strictly speaking, a compound is any statement that has greater logical form than an atomic. So all the following are compounds:

Ford and Chevy make muscle cars.

That forest has oak trees and is home to many woodland critters.

I did not go to the store.

If we go to the movies, we'll get some popcorn or nachos.

The savvy student will note that some of these statements have more than one subject. Some have more than one predicate. The third example may seem like it does not belong on the list, because there is one subject, and we say one thing about it. However, a closer look reveals that what we say about that one subject is more complex in its logical form than what we see in atomics. Let's break it down:

Subject: "*I…*"

Predicate: "*…did not go to the store*."

Question: Well, where *did I* go then?

Notice that you cannot answer this question. The statement does not *declare* where I actually went. This statement tells us nothing about my real activity—it makes no attempt to assert something definitive. This is very different from an atomic statement, which we defined as a simple declarative statement. The atomic statement "*I went to the store*" does quite definitively describe where I went (once we clarify or denote which store). If I ask my question again: Where did I go? This statement provides us with a substantive answer.

However, our statement "*I did not go to the store*" does something very different—it denies a declaration of my trip's destination. In doing so the speaker has taken a definite statement *and then* said "no" to the truth of that statement. In terms of the logical structure of their statement, that's *more complex* than an atomic. Their statement is a compound—in this case, not in the sense of combining more than one atomic, but of still *doing more* with one atomic than simply asserting it.

> Put simply: atomic statements always make a single *affirmative assertion* about a single subject. If you are doing more than that, you are making a compound statement.

When we make compound statements, we are performing logical operations on smaller statements. This is what we do. We are not content to think, reason, believe, and communicate things about the world in the simplistic manner of atomic statements. We see important relationships in the world and need a way to express our understanding of these. Enter: logical operators.

Logical Operators

Strictly speaking, we should introduce these as *truth-functional* logical operators. The reason is that when we do things with smaller statements, we sometimes do them in a way that has a direct impact on the truth of the larger statements we make. Put differently, when a larger statement has been made by using these logical operators, the truth of these larger statements can be determined by the truth value of the smaller statements. Truth-functional operators take in the truth value of the component statements and spit out a new truth value depending on the nature of the logical operation being performed. Consider these intuitive examples and see if you can sense when the following examples are true and when they are false:

I went to the store, and *I went to the bank.*

I went to the store, or *I went to the bank*.

Our speaker is talking about the exact same thing (themselves), and even describing that thing in the same ways. Yet we see that these two statements are very different in their truth values. The second statement may be true at the exact same time that the first statement is false. This is because the two compound statements *were made* in a different way that bears directly on the truth value of the whole statement. The compounds here were both made with truth-functional logical operators.

Logical operators are also sometimes called "statement *connectives*" (or "truth-functional *connectives*") because of the role they play in constructing larger compound statements. They act like a kind of glue to bond together atomic statements (or, to again play off of our physical analogy, these are *the bonds* of our molecules). However, the first of these (which we just saw above) does not really bond together statements, it simply performs a logical operation on a statement—it denies it. We will review this as well as four other logical operators to help flesh out our symbolic language.

Table of Logical Operators

Operator Name	**Common Name**	**Symbol**	**Symbol Name**
Negation	"not"	~	(tilde)
Conjunction	"and"	·	(dot)
Disjunction	"or"	v	(vee)
Material Conditional	"if, then"	→	(arrow)
Material Biconditional	"if and only if"	↔	(double arrow)

In our text, we will use these symbols to represent the important logical operations speakers make when constructing their compound statements. You will find some variation in the symbols used in other logic textbooks. You will likely also find the same (or similar) operators in a math or a computer programming textbook, and these often use different symbols. If you use those symbols to translate, your math teacher will be oh so very proud. Your logic teacher, on the other hand, will wonder why you are translating into a different language—much like your German teacher *does not care* if you are familiar with negation as it is expressed in the Russian language ("nyet")—they want to see that you can translate the English "no" into what is proper for *that* class ("nein").

Negation

Our previous example helps us recognize the basic function of the negation:

I did not go to the store.

Negations are denials of other statements. Our translation scheme for these is rather easy. We simply insert a "~" in front of whatever statement we want to deny. In this case, someone wants to claim:

I went to the store.

Translation: S

I deny that this is true. I say:

I did not go to the store.

Translation:~ S

We can always deny a statement. We're not always right when we do, but in principle the negation is a simple function. The only real question is whether we are denying a simple atomic statement (as above) or whether we are denying a compound statement. After all, my friend may predict:

If you go to the store, you will get some milk.

Translation (we'll see this in a bit):

S → M

I am always at liberty to reject this claim. In doing so, I should be careful that I do not deny some other claim—I want to deny *my friend's* claim. So I will *not* say:

I won't go to the store.{ translated as ~ S }

Nor will I say:

I won't get some milk.{ translated as ~ M }

Those would not deny what my friend said, for they never said I *would* go to the store, nor did they say I *would* get some milk. They said IF I go to the store, THEN I'll get some milk. They expressed a relationship between these two claims. So, I need to deny *that* relationship. I will not deny the individual statements, nor express *the same relationship* my friend expressed but with each part denied { ~ S → ~ M }. That will not do either. I will deny the relationship itself.

To do so, I will likely say something like:

It is not true that if I go to the store, then I will get some milk.

Translation: ~ (S → M)

Conjunction

Conjunctions are unions of two statements. The speaker wishes to assert two things within a single sentence.

The two "things" may be about two separate subjects, or they may be two different descriptions about a single subject. For example:

Brandon and Lily go to college.

Brandon goes to college and studies economics.

Both of these are conjunctions. We may find it easier to paraphrase compounds such as this to make the two statements clear. For example:

Brandon goes to college, and Lily goes to college.

Brandon goes to college, and he studies economics.

Now each part of the conjunction is fully and explicitly expressed. Later this will help us to translate these statements. This is especially true when we come across larger conjunctions whose parts are themselves compound statements. In these examples, our translation scheme would be (respectively):

B · L

B · E

For now, let's just note that we will often need to talk about the two components of a conjunction. We call each part a "**conjunct**" and distinguish them by left side/right side. Only conjunctions have conjuncts.

Disjunction

We use disjunctions to express alternatives between two statements. The most common way is to use "or" to divide them up. As with all other compounds, these two options can be regarding a single subject, or they can be options between two different subjects (or a combination of these). Consider:

Mark will plant the garden or he will sleep in.

Mark or Tom will plant the garden.

Mark will plant the garden alone, or Tom will sleep in.

The translation schemes for these would be as follows:

M v S

M v T

A v B

The savvy student will note that we had to take care not to confuse our atomics. The claim that "Mark will plant the garden alone" is a different claim than simply "Mark will plant the garden." Thus, they require different translations into different capital letters. Likewise, the claim "Mark will sleep in" is clearly a different statement than "Tom will sleep in." So they too required different translations.

As we will see below, there are several other ways to express disjunctions. What matters now is that we need to understand that the logical operator of disjunction expresses a very liberal view of these options. The best way to express it in English is likely:

At least Mark or Tom will plant the garden.

The reason this helps is that it reveals to us that we are making a very, very weak statement. The logical disjunction is a noncommittal statement. The speaker is not going very far out on a ledge to assert much of anything. They simply want to establish the options—*but hey, don't hold me to that*—I mean, those options might even include the one where both Mark and Tom do the gardening. This is why we call the logical disjunction an "**Inclusive-Or**" statement. A logical disjunction *includes* the possibility that both sides of the disjunction come out true…it also includes the possibility that only the left side is true…it also includes the possibility that only the right side is true. See: very weak, very noncommittal.

We call each side of the disjunction a "**disjunct**," and as with the conjunction, they are distinguished by left side/right side. Only disjunctions have disjuncts.

Material Conditional

Consider the following example:

"If Lincoln wins the election, then Lincoln will be president."

We could treat this like an atomic sentence, but then we would lose a great deal of important information. For example, the sentence tells us something about the relationship between the atomic statements "Lincoln wins the election" and "Lincoln will be president." To make these relationships explicit, we will have to understand what "if…then…" means. Thus, it would be useful if our logical language were able to express these kinds of sentences in a way that makes these elements explicit. Let us start.

The sentence "If Lincoln wins the election, then Lincoln will be president" contains two atomic sentences: "Lincoln wins the election" and "Lincoln will be president." We could thus represent this sentence by letting

"Lincoln wins the election"

be represented in our logical language by

P

And by letting

"Lincoln will be president"

be represented by

Q

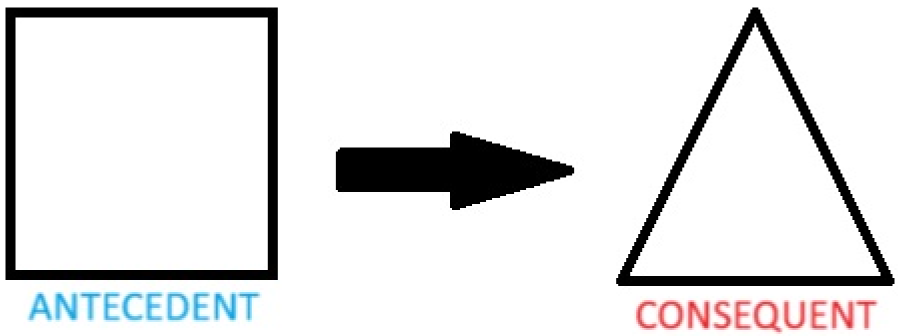
Then, the whole expression could be represented by writing

If P then Q

It will be useful, however, to replace the English phrase "if…then…" with a single symbol in our language. Our symbol is "→." Thus, we would write

P→Q

This kind of sentence is called a "material conditional." It is also often simply called a "conditional." The first constituent sentence (the one before the arrow, which in this example is "P") is called the "**antecedent**." The second sentence (the one after the arrow, which in this example is "Q") is called the "**consequent**."



One last thing needs to be observed, however. We might want to combine this complex sentence with other sentences. In that case, we need a way to identify that this is a single sentence when it is combined with other sentences. There are several ways to do this, but the most familiar (although not the most elegant) is to use parentheses. Thus, we will write our expression

$(P \rightarrow Q)$

This is useful when we want to say something like:

"*I think Douglas will win easily, but admit that if Lincoln wins the election then Lincoln will be president*"

If we let "I think Douglas will win easily" be translated as

D

Then the whole expression translates as

$D \cdot (P \rightarrow Q)$

Note how the parentheses help our translation stay true to the original English expression. As we will see shortly, without these in place, our symbolic translation could be interpreted in a very different manner than what our original English speaker intended.

Material Biconditional

Strictly speaking, we don't really need the logical operator that does the work of the material biconditional. We could get by translating what people say using a combination of and, or, and if-then quite easily. The trouble is, this may be *too easy* in many cases. We may be inclined to mistranslate what people intend if we don't have something to reference that would make their original intent clear. Enter: the **Material Biconditional**.

Biconditional statements are what most people actually hear when they hear a conditional statement. They hear a very tight relationship between what is said on one side and what is said on the other side. For example:

If you do the dishes, then I'll take you to get some ice cream.

A proper translation would be:

D → I

People who hear this often hear that each side is somehow dependent upon the other. This is false. So, not surprisingly, this usually leads to error and confusion. Therefore, we find it is quite helpful to have a way to explicitly express (and distinguish) the relationship of a conditional statement (→) from the tighter relationship expressed in the biconditional. The name is different, the symbol is different (↔), and the conditions of truth are different. All of this is helpful in communicating clearly what we intend.

I'll take you to get some ice cream *if and only if* you do the dishes.

That's what most people heard when they read the first statement. And if that is what was intended, the original speaker misspoke and would do well to take a logic class. This would help everyone else focus in on what relationship was intended in the speaker's claim. We could easily translate this as:

I ↔ D

On the other hand, if the original speaker really did mean to express a single material conditional, then the rest of us would do well to listen more carefully to what they were saying and not put words (or meaning) in their mouth. Our translation would not make use of the double arrow.

There is no special name for each side of the material biconditional.

GRAMMAR—WELL-FORMED FORMULAS

We have already introduced the concept of syntax—the form of a statement. All languages have rules for syntax, and we typically think of this as their grammar rules. These are standards to ensure that the form of the expressions we make are consistent for all users and that the statements we make clearly communicate our desired meaning.

In our symbolic language, we need rules to make sure that the statements we express unambiguously reflect the meaning of those we are translating. Fortunately, there are only a small number of these rules. We refer to them as the conditions of "well-formed formulas," or "WFFs" (pronounced "wiffs") for short.

Statement Variables

To learn these rules we need to introduce another concept: the statement variable. A statement variable is (like all other variables) just a placeholder for a statement. Unlike atomic statements, statement variables do not actually say anything—they just act like a marker for a statement of some kind that does say something. What is most important is that we understand that these placeholders **do not presuppose what kind of statement they represent:** they could stand for an atomic or *any* kind of compound statement. They are simply blank spots, and they help us talk about the general form of a statement. We will use the following to represent statement variables:

☐ (box)

∆ (triangle)

On rare occasion we will have cause to use two other statement variables:

◯ (circle)

☆ (star)

In all cases, it should be clear that these are just symbols that represent the presence of some kind of symbolic statement. None of them have any priority nor preference; they do not exclude one another, nor insist upon difference (i.e., the statement that we eventually find inside a box can be the exact same statement we eventually find inside a triangle, etc.). They are simply *empty spots* in which a symbolic statement will go.

Many logic textbooks take their cues from math textbooks and use letters for statement variables. So you will see this:

"*p*," "*q*"

Or even worse:

P, *Q*

You might also see Greek letters used:

ϕ, ψ

These are *exceedingly* unfortunate (the Greek is tolerable; the rest are counterproductive). We have just learned that "capital *letters*" are used in our symbolic language to represent atomic statements. So why muddy the waters with *more* letters?

These ways of symbolizing statement variables dupe many students into thinking you are only allowed to insert atomic statements in these spots—this is NOT true. Unfortunately, the rest of the logic world has not caught on to this troublesome conflation. So let's set the record straight:

A statement variable can contain ***any*** statement.

We will use boxes and triangles, because after all, *anything can fit inside a box* (you just need a big enough box).

There are only six grammar rules in our symbolic language. They are as follows:

Six Conditions for Well-Formed Formulas

1. All atomic statements are WFFs
2. ~ □ is a WFF, if and only if □ is a WFF
3. (□ · △) is a WFF, if and only if □ and △ are WFFs
4. (□ v △) is a WFF, if and only if □ and △ are WFFs
5. (□ → △) is a WFF, if and only if □ and △ are WFFs
6. (□ ↔ △) is a WFF, if and only if □ and △ are WFFs

Note that we use the ( ) to begin and end our grammatically correct expressions in our symbolic language. This is much like the English rule of starting every sentence with a capital letter and ending with an appropriate punctuation (generally a period when a statement is made). We are well-served when we know clearly when an expression begins and when it ends.

We will make use of two conventions that help us clean up the visual appearance of larger compounds. These two are as follows:

1. Drop the outermost parentheses (you must always retain the inner parentheses)
2. Alternate ( ) with [ ] to group smaller compounds in a larger compound

For example, strictly speaking, the following is a WFF:

(A v B)

This statement conforms to the 4th condition for WFFs. We see a statement that perfectly conforms to the pattern:

(☐ v Δ)

We read this as: open parenthesis, A, v, B, closed parenthesis. When we look inside the ☐ and the Δ positions we see atomic statements (which are WFFs by the 1st condition for WFFs). This is all well and good, but let's face it, nobody is going to be confused about the meaning of our statement if we just say:

A v B

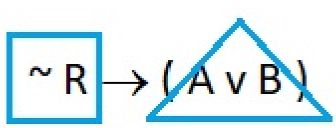
That's perfectly clear, even though we dropped the *outermost* parentheses (and in this case, the only parentheses). However, if our statement were more robust, we would quickly find our statements are not clear if we start dropping all the parentheses. For example, I may want to say:

(~ R → (A v B))

We can clean this up by dropping the outermost parentheses:

~ R → (A v B)

Again, there is no risk of confusion in doing this. We can readily tell what the box is and what the triangle is in this case.



However, consider this formula:

~ R → A v B

What do I mean by this formula? I could mean:

(~ R → A) v B

I could also mean:

~ (R → (A v B))

Or I could mean:

~ (R → A) v B

All of these variations mean something completely different. Once we drop all the parentheses, our symbolic statement becomes highly ambiguous. *This is not good*. We do not run this risk if we are just dropping parentheses that would belong on the outside of our statement's main components. However, anything *INSIDE* a box or a triangle needs to retain the parenthesis if it is a compound statement that requires them in the Six Conditions for WFFs.

Just to clarify:

Atomic statements do not require any parentheses.

Negations do not require any parentheses UNLESS what is inside their box is a compound

This last point deserves an illustration. So consider:

~ M v L

This is a WFF by our first convention. It satisfies the 4^{th} condition for WFFs.

However, the following statement does not satisfy the 4^{th} condition; it satisfies the 2^{nd} condition:

~ (M v L)

Here we have a negation: a statement that satisfies the 2^{nd} condition for WFFs. What is *inside* the box is a disjunction—it is inside the box of ~ □, so it must retain the full account of the 4^{th} condition for WFFs. This means that *it must retain the parentheses*.

On the other hand, look again at our first example:

~ M v L

Notice how we have a statement that conforms to (□ v Δ) without the *outermost* parenthesis.

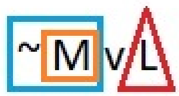
Condition 4 tells us that this is fine so long as what is inside the □ and Δ are also WFFs. Inside the □ is the statement:

~ M

This conforms to the 2^{nd} condition for WFFs:]

~ □ is a WFF, if and only if □ is a WFF

Now when we look inside this little □, we find an atomic statement.



The 1st condition for WFFs tells us that atomics are WFFs all by themselves—*there should NO parentheses around it*. So, no parentheses around a negation and no parentheses around an atomic.

Last, a quick word on our second convention that allows us to alternate ( ) with [ ] (we call the latter "brackets"). This just helps us see the groups of compound statements when statements get very large. Consider the following:

~ [(N → A) v ([M v K] → U)]

The brackets simply help us recognize the fact that this is a giant negation (and more precisely, this is a negation of an Or statement:

~ [(N → A) **v** ([M v K] → U) **]**

For most folks, things can get blurry when we only use parentheses:

~ ((N → A) v ((M v K) → U))

The brackets help a bit.

There is no hierarchy between ( ) and [ ] as there is with English punctuation (commas are subordinate to semicolons). So the translator is at liberty to use them as they see fit, so long as we are not mix and matching the opening and closing punctuation (i.e., if you open with a bracket, you must close with a bracket, etc.).

The Main Connective

We should note that the second half of each condition for WFFs stipulates that the component parts of a compound statement must themselves be grammatically correct. This is extremely important and helpful. In larger compounds we may find it difficult to tell what kind of statement is being asserted. Knowing that the grammar rules require each part of a large compound to be grammatically correct means we have a way to test a statement to see if it is this or that kind of compound. Let's look at the last example:

~ [(N → A) v ([M v K] → U)]

This is a challenging statement to identify at first. We may not know what kind of statement it is, but the WFF rules remind us that however it is carved up, each component must be a WFF. So we can test it.

Say our gut tells us that this is an "or" statement (a negation). If so, then there are a □ and a ∆ that must themselves be WFFs.

To the left of the "v" we see this:

~ [(N → A)

To the right of the "v" we see this:

([M v K] → U)]

Are either of these WFFs? No. So our gut misled us, and we know that this is not an "or" statement. In fact, if we thought we could use the other "v" to identify this as a grammatically correct statement, the situation becomes worse. To the right of *that* "v" we see:

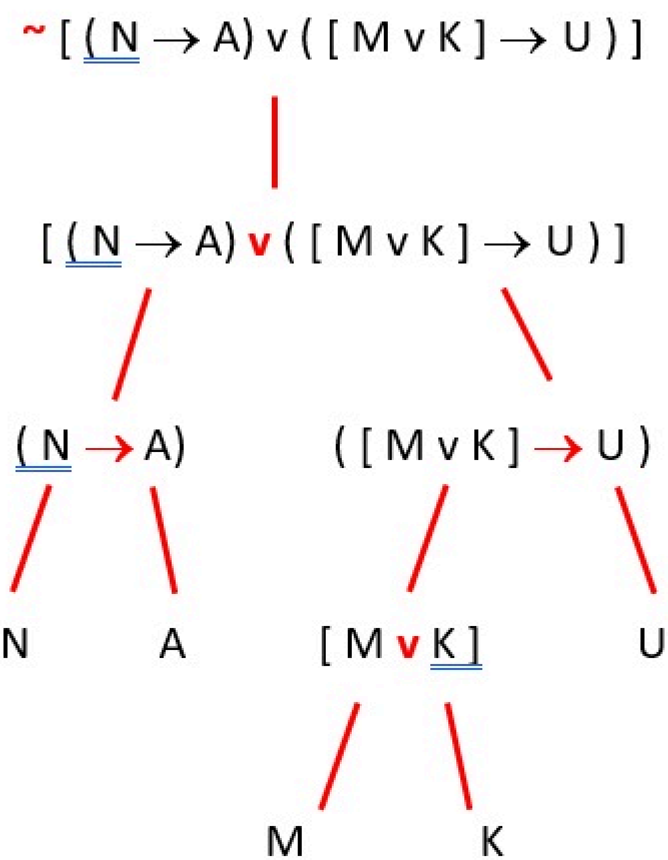
K] → U)]

This is clearly not a WFF.

Keep going: try every operator up there and you will find only one that produces a ☐ or a △ that is a WFF. This turns out to be the "~" operator. Like all ~ ☐ statements, to the right of the "~" must be a WFF. In this case we see:

[(N → A) v ([M v K] → U)]

Note that we are inside a ☐ now, so the outer brackets must be retained. This makes the statement here conform to the 4th condition for WFFs. We could continue to break this statement down into each of its component parts until we arrived at the smallest units, the atomics. Like this:



You can do this exercise with any size compound statement. Indeed, making a tree diagram like this is helpful practice to train your eyes to start identifying the main connective of every statement. Doing so will vastly improve your ability to perform on just about everything we do in logic from this point forward.

The **main connective** (the **MC**) is one of the single most important concepts you will need to master to pass a logic class. We can understand the MC as having three defining features:

1. **The MC is the last logical operator used to construct a statement**
   (It is the last in working your way UP from atomics to the final statement at the top of your tree diagrams.)
2. **The MC is the bearer of truth for a statement**
   (We will see this later in our work on truth tables.)
3. **The MC is the identity of the statement**
   (We see this now in how statements conform to one of the conditions for WFFs, but we will make heavy use of this feature when we do logical derivations or logical proofs.)

Practice making tree diagrams until you get very good at just looking at a statement and identifying what kind of operator is the Main Connective.

Linguistic Variants

There are many ways to express logical operators in a natural language. We call these "alternate phrasings" or "linguistic variants of the logical operators."

Since we are working in English, we will look at some of the most common ways the English language captures these logical operations. One general word of warning: like most natural languages, English is very robust and allows speakers to use the same terms in different ways. Put differently, not every instance of a linguistic variant is being used as a logical operator. For example:

Tom and Jerry were great characters.

Tom and Jerry was my favorite cartoon.

In the first statement the "and" is being used as a logical operator. We would do well to paraphrase that statement as:

Tom was a great character, and Jerry was a great character.

In the second statement ("*Tom and Jerry* was my favorite cartoon"), this is not how the speaker is using the "and" in their statement, for it is simply a part of the name of that one show. In other words, we would *not* want to paraphrase the second sentence as:

Tom was my favorite cartoon, and Jerry was my favorite cartoon.

Heck, there wasn't even a cartoon named "Jerry," so this would be a terrible paraphrase and result in an incorrect translation. We'll try to note some of these non-logical uses of the linguistic variants along the way.

Negation

Denials are almost always intended, so we rarely have to worry about a speaker using an English variant of the negation in any way other than to actually mean "No, that claim is not true." Common variants are:

"…not ☐"

"…n't ☐"

It is false that ☐

☐ is not true.

The claim ☐ is wrong.

☐ is not so.

☐ can't be the case.

I don't believe ☐.

Despite what we just said about not having to worry about variations in meaning, we can see a few cases here where a speaker may be expressing a subtle claim. For example, "I don't believe ☐" usually means ~ ☐. However, the speaker may be very cagey and only intend to say something about themselves. They may not intend to deny the truth of what was said *about the world*, but rather deny the content of *their own mind*. Cagey.

Additionally, the phrase "☐ can't be the case" is usually meant in a colloquial voice to mean ~ ☐. We may blurt this out when we are surprised or were convinced we would discover something different. However, if we strictly adhere to the phrase, it means something stronger than "☐ *is* not true"—it means "☐ *cannot* be true" (it is *impossible*). This is a much stronger denial than what we typically mean when we say something like "The lights are not on"—I only mean they are not on at the moment, I don't mean it is impossible for them to be on. For purposes of translation, we will get by with the casual meaning of these variations and go with a simple schematic of ~ ☐.

CAUTION

One word of caution regarding negations is in order. Many people conflate **conceptual opposites** with logical negations. Some ideas have counterpoints that we associate with them. For example, I say "dog" to you and ask, what's the opposite? You might say the opposite of "dog" is "cat." We hold dogs/cats as some kind of conceptual pair. Of course, there is nothing denied in "cat" that is expressed in "dog" (and cat/dog is only a conceptual pair in some countries). Still, people may act as though they are opposites. For example:

Mary: I had to rush my baby to the vet yesterday!

Peter: Oh no! What's wrong with your dog?

Mary: Huh, I don't have a dog.

Peter: Oh sorry, what's wrong with your cat?

Mary: I don't have a cat either—I have a snake.

We do this more frequently with tightly packed concepts. For example:

Howard: Darn, the team I bet on didn't win yesterday.

Yuri: Sorry to hear they lost.

Howard: What? They didn't lose, the game was a tie.

Winning and losing are conceptual opposites. They are not logical denials of one another. Howard spoke the truth when he said his team did not win. Yuri made the mistake of thinking he could "translate" that statement as:

L: Howard's team lost

When he should have simply translated it as:

~ W (where W: Howard's team won.)

Many people make such errors regarding logical negations. In translating, it is better to simply put a tilde in front of whatever the claim was rather than try to "think through" *what we believe* the speaker's meaning was in offering the negation.

Last word of caution: beware the "n't" hidden in many statements. If I say, "I didn't go to the store," we may find ourselves hiding the logical operation in this statement by translating it as an atomic (e.g., S). This would hide useful information regarding what was actually said. This tiny "n't" is a logically significant move made by the speaker. We should strive to capture it faithfully in our translation.

Conjunction

The logical conjunction is well-captured with the English "and" in most cases. However, there is a range of other English terms that can be used to capture the basic form of ☐ · ∆. The following are the most common English variants:

☐, and ∆

☐, yet ∆

☐, as well as ∆

☐, in addition to ∆

☐, but ∆

☐, however ∆

☐, nevertheless ∆

☐, even though ∆

Two things need to be addressed: rhetoric and a discomfort.

First, many English expressions use these terms for rhetorical effect. For example, I may start a statement with one of these: "However, we should also consider waiting for more information." Strictly speaking, the "However" here is not used as a logical conjunction. This is just a smooth way to continue to the next statement. Oddly, for years English teachers instructed students to never begin their sentences with "But" while simultaneously encouraging many of the other openers as rhetorically viable options (e.g., "Yet, we might find it equally wise to contact the source for their input."). To logicians, these are all the same—*none of the openers* are treated as logical operators.

Second, when we review our list of variants, we immediately see that some of these English terms contain a sense of contrast or denial that is not contained in the logical conjunction. So, if you have some reservations about translating "yet" or "but" as a conjunction, you're not alone. Terms such as "nevertheless" and others often give the audience the sense of opposition between the two component statements. This does not come across in the logical conjunction; this sense of the original English is lost in translation. When we look at a statement such as:

I went to the store to buy milk, but orange juice was on sale.

Our translation would be:

S · O

In our original English we never said "no" to anything. We affirmed that we actually went to the store. We also affirmed that orange juice was on sale. Both statements are affirmed. The sense of surprise or possible *insinuation* (not actually stated) that we did not get milk is lost in translation.

CAUTION

Be careful with this important variant:

Both □, and △

The addition of "both" should be treated as more than fluff. This is a useful way for our speaker to alert the audience that a conjunction is being formed. This is important when the speaker is trying to group two statements within the context of *another* logical operation. Consider the following:

Both Mike and Tom did not win.

Mike and Tom did not both win.

These two statements mean different things. They are true under different circumstances. And our author is trying *to help us* recognize their intent. We need to attend carefully to listen well. So which translation do you think belongs with which statement?

~ □ · ~ △ ~ (□ · △)

Take a moment to listen carefully to how the author uses "both" in their two statements. The term "both" is like a flag; it alerts the audience that a dot is forthcoming. The question is, what is the nature of each conjunct?

In the first statement, we are alerted right away that two things are forthcoming: "Both □, and △."

Inside the □ (as well as in the △), we find the claims about each not winning. We could paraphrase:

Both Mike did not win, and Tom did not win.

Thus, our translation schematic would follow the form:

~ □ · ~ △

Yielding the actual translation:

~ M · ~ T

In the second claim (*Mike and Tom did not both win*), the "both" again flags a grouping: "...both win." However, in the English, this group is preceded by a denial of that grouping. A rough paraphrase would be:

It is false that both win.

Thus, our translation schematic would follow the form:

~ (□ · △)

Providing a full paraphrase makes the original English clear:

It is false that both Mike won and Tom won.

Our translation is now easy:

~ (M · T)

TIP: On Flags and Negations

Another way to look at this is to follow a simple rule that usually helps:

Negations negate what they immediately encounter

When we look at the English statements, we find an English "not" that runs right into something:

Both Mike and Tom did not win.

Mike and Tom did not both win.

In the first sentence, the English "not" immediately encounters the claim about winning. Which claim? Both claims. The one about Mike and the one about Tom. Thus:

~ M · ~ T

In the second sentence, the *English "not" immediately encounters the English flag "both,"* so we translate this as *a tilde that immediately encounters an open parenthesis* that contains this type of grouping.

~ (M · T)

As we will see, other logical operators can be flagged by English words, and this same rule applies. Pay close attention to what the "not" or the ~ immediately encounter.

Disjunction

We have seen that the logical disjunction is, in a way, the opposite of a conjunction. Whereas the conjunction allows a speaker to affirm two things, the disjunction allows them to merely suggest that at least one of two options is true. There are several ways to express this relationship in English. For example:

☐, or △

At least ☐, or △

☐, unless △

Either ☐, or △

☐, and/or △

The savvy student probably already notices that some of the variants include "flags" for the logical disjunction. After all, if I open my statement "Either…yaddah, yaddah, yaddah…." you (the audience) know that sooner or later I have to say "OR" and then continue my yaddahs. The "either" works well this way. The "at least" can also be used this way, but you will find some speakers use it excessively as a rhetorical device. For example,

At least clean your room once in a while.

This "at least" is not a flag for a disjunction. However, you will never see "either" used rhetorically.

Inclusive vs Exclusive Or

We already noted that the logical disjunction is an Inclusive Or. In English, the best way to properly capture this logical relationship is found in:

At least ☐, or △

Or stronger still: *At the very least* ☐, or △

☐, and/or △

While I find some speakers prone to rhetorically use "and/or" for effect, this is entirely the fault of those speakers. The English "and/or" very clearly and explicitly expresses an Inclusive Or.

If we can ignore rhetorical flourish for a bit, we should consider that many speakers (and audiences) expect their English "or" to be received differently than what is meant by the logical disjunction. After all, if I say, "We'll go to the movies or to the carnival tonight," you probably think we'll do only one or the other. Indeed, a better paraphrase for our sample statement here would be:

We'll go to the movies or to the carnival, but not both.

This expresses the removal of one of the included options of the Inclusive Or. We call these "Exclusive Or" statements. They can be captured in our symbolic language by carefully translating our paraphrase.

We'll go to the movies or to the carnival, **but** not both.

We should see that we have two things affirmed: here are the options, AND here is the denial of one of the included options we would find in a standard disjunction. So, this translates as:

$(M \vee C) \cdot \sim (M \cdot C)$

The main connective of this statement is the blue dot. What this reveals is that when a speaker intends to communicate an Exclusive Or, they are actually expressing a conjunction.

The Special Case of UNLESS

Some people are very frustrated with treating the term "unless" as a disjunction. Some just don't like the term "unless" at all. Still, it causes enough trouble that we should address it here. There's no shortage of people who hear a conditional statement when they hear "unless." This is fine, so long as we know how to properly translate that conditional statement. Most folks get this wrong, because it takes quite a bit of thought and effort. Put differently, it can be tricky to correctly translate "unless" as a conditional statement.

Conversely, translating "unless" as a disjunction is *very* easy! All you have to do is put a "v" on top of the word "unless" and translate the left / translate the right / and done! Very easy, very little extra thought put into it. Here's an example:

Either we get those reinforcements, or the fort will fall tonight!

Let's think this through. We might translate this as a straight conditional statement:

$R \rightarrow F$

However, if we read that back we get:

If we get those reinforcements, then the fort will fall tonight.

That's clearly wrong. The speaker did not mean that; they meant something different. Something like this:

If we do not get those reinforcements, then the fort will fall tonight.

Now we can correctly translate their original statement into the proper conditional:

~ R → F

Here's another way to do it. Remember the original statement:

Either we get those reinforcements, or the fort will fall tonight!

We can do as I suggest, and just slap a "v" over the "or" term:

Either we get those reinforcements, **v** the fort will fall tonight!

Translation:

R **v** F

Done.

The Special Case of NEITHER, NOR

Often, we hear people use the phrase "neither nor" to indicate a special relationship between the two smaller statements. We can properly translate this in two distinct ways. So, let's try a test:

Neither Porsche nor Tesla make cars with a true American spirit.

What did *you* hear? See, most people probably heard our speaker make two denials. They hear:

Porsche does not make cars with a true American Spirit, **and** Tesla does not as well.

That's something like what most people heard. So we can successfully translate this as a conjunction of the two denials:

~ P · ~ T

Our speaker wants to deny both claims. That's a good translation. However, some people (in my experience, they are the minority) hear something different. They hear an "either, or" statement…*that was denied*.

*N***either** Porsche *n***or** Tesla make cars with a true American spirit.

These folks are also correct! So, we can successfully translate the same statement as the Negation of an Or statement:

~ (P v T)

Advice: Whatever *you hear* when you hear "neither, nor" is what you should use. Both translation schemes are logically equivalent to one another and express the exact same relationship between the two component statements. When we look at truth tables, we will be able to prove this clearly.

Material Conditional

English includes many alternative phrasings that appear to be equivalent to the material conditional. Furthermore, in English and other natural languages, *the order of the conditional will sometimes be reversed*.

This permits speakers to vary their conditional statements with a bit of personal style and speaking preference while keeping the same meaning.

We can capture the general sense of these cases by recognizing that each of the following phrasings would be translated in the form of □→Δ. The most important thing to notice is *the position* of the antecedent in both the original English and the symbolic translation.

(In these examples, we mix English and our propositional logic, in order to illustrate the variations succinctly.)

If □, then Δ.

Δ, if □.

On the condition that □, Δ.

Δ, on the condition that □.

Given that □, Δ.

Δ, given that □.

Provided that □, Δ.

Δ, provided that □.

When □, then Δ.

Δ, when □.

□ implies Δ.

Δ is implied by □.

□ is sufficient for Δ.

Δ is necessary for □.

Δ is a prerequisite for □.

□ guarantees that Δ.

The Special Case of *Only*

An oddity of English is that the word "only" changes the meaning of "if." You can see this if you consider the following two sentences.

Fifi is a cat, if Fifi is a mammal.

Fifi is a cat only if Fifi is a mammal.

Suppose we know Fifi is an organism, but we don't know what kind of organism Fifi is. Fifi could be a dog, a cat, a gray whale, a ladybug, a sponge. It seems clear that the first sentence is not necessarily true. If Fifi is a gray whale, for example, then it is true that Fifi is a mammal, but false that Fifi is a cat; so, the first sentence would be false. But the second sentence looks like it must be true (given what you and I know about cats and mammals).

We should thus be careful to recognize that "only if" does not mean the same thing as "if." (If it did, these two sentences would have the same truth value in all situations.) In fact, it seems that "only if" can best be expressed by a conditional where the "only if" appears before the consequent (remember, the consequent is the second part of the conditional—the part that the arrow points at). Thus, sentences of this form:

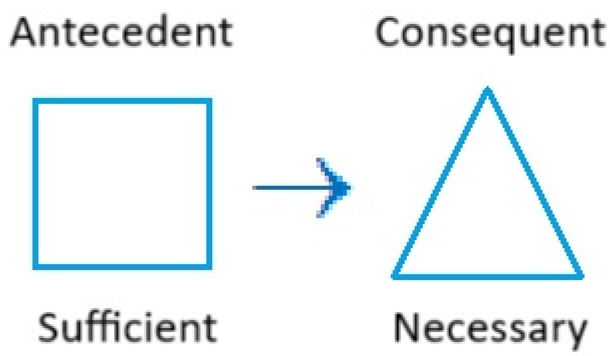
☐ only if Δ.

Only if Δ, ☐.

are best expressed by the formula

☐ → Δ

The special case of "only if" as well as its synonyms "necessary condition" (as well as variants, "…is needed" or "…need to" etc.) can be tricky to think through. Distinguishing between "sufficient" and "necessary" conditions may also pose difficulty. So, we will find it helpful when translating to pay attention to the positions of antecedent and consequent. Just drop into those spots whatever was claimed as necessary or sufficient. Memorizing this diagram may help:



Simply put:

If something is claimed to be "sufficient," then stick it in the antecedent.

If something is claimed to be "necessary," then stick it in the consequent.

For example:

You need to study to do well in class.

Q: What was asserted as the thing that was needed? A: "You study."

So just stick that claim in the consequent position (i.e., the Δ spot).

W → S

When we read that back, we get something intuitive like: "If you did well in class, then you studied." We can confidently assert this, because "you study" is a necessary condition for doing well. We would not want to say, "If you study, then you do well" (translation: S → W), because that's false (studying is not sufficient to do well—you know, you have to also take the exams, come to class, turn in your homework on time…). More importantly, we would not want to translate our original (true) claim into something that the original speaker never said. *This would be a failure of ours to listen well*.

Material Biconditional

We previously noted that we don't really need the biconditional operator. It exists to give us a useful way to communicate with precision and avoid confusion in the ways we express conditional relationships. In English we note this with how many more words it takes to express a material biconditional (and there are really only two):

□ if and only if Δ.

□ is a necessary and sufficient condition for Δ.

That's a lot of words. Contrast these with how briefly and simply we can express the simple conditional. So, there's a tip: if there are many more words coming out of that person's mouth, they mean a biconditional. Simple.

Last, note that when translating from English into our symbolic language, we put the component statements in the order in which they appear in the original English. This too is very different from our translation technique when addressing a simple conditional, where the English typically flags the antecedent regardless of the order in which it was said (e.g., "I'll get some milk, **if** I go to the store."). However, with the biconditional we just translate in the order of the original English expression. Again, simple.

6.

TRUTH TABLES

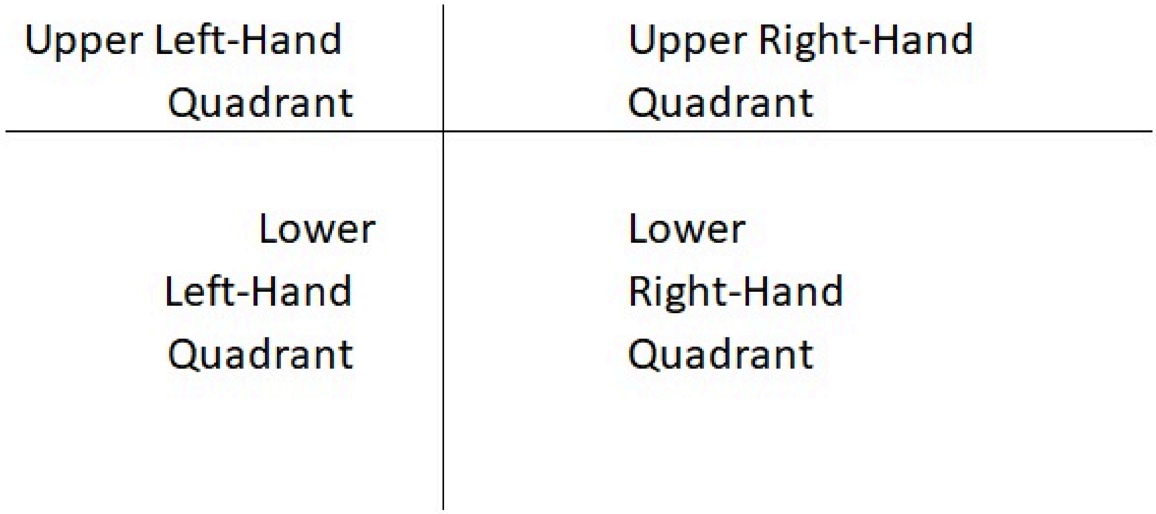
We have seen how challenging it can be to evaluate arguments for their informal qualities. When we turn to evaluating their formal qualities, there are many tools at the logician's disposal to ease the process. Truth tables are nothing more than a way to visualize the core concepts and formal properties of statements. As such, they also help us quickly evaluate arguments. In this sense, truth tables are nothing more than pictures we draw. We want a way to draw the concepts surrounding truth and possibility. So, if a picture is worth a thousand words, then a truth table is worth *all possible worlds*…

Given our commitment to bivalence, we know that there are only two possible truth values for any given statement, P: it can be either true or false. We have a way of writing this that is called a "truth table." For an atomic sentence, the truth table is trivial, but when we look at other kinds of sentences, their truth tables will be more complex.

The idea of a truth table is to describe the conditions in which a sentence is true or false. We do this by identifying all the atomic sentences that compose that sentence. Remember that larger statements are composed of truth-functional logical operators—so the truth of any compound will depend on the truth value of the smaller atomics out of which the compound is made.

Setting Up a Truth Table

The set up for a truth table is simple. We will only have four quadrants to our table, so only two lines need to be used to organize the information we need. This will look like this:



You might even think of truth tables as being divided in two halves. On the left is the basic information we need, and on the right is the statement(s) we want to evaluate and the information about it.

So, on the left side, we stipulate all the possible truth values of these atomic sentences and write these out. On the right side, we then identify under what conditions the sentence (that is composed of the other atomic sentences) is true or false.

The idea is that the sentence on the right is dependent on the sentence(s) on the left. So the truth table is filled in like this:

Atomic sentence(s) that compose the dependent sentence on the right	Dependent sentence composed of the atomic sentences on the left
All possible combinations of truth values of the composing atomic sentences	Resulting truth values for each possible combination of truth values of the composing atomic sentences

One important note before we go any further. When we say we will fill in the upper left quadrant with all the atomic statements that compose our statements on the right, we must follow one simple rule: list the atomics in the SAME order in which they appear in the upper right (when you read from left to right). We do not put them in alphabetic order, and we do not spell out funny words with them. We always put them in *exactly the same order* in which they appear in the upper right (skipping any repeats).

We stipulate all the possible truth values on the bottom left because all of the logical properties that we learned are defined in terms of different possibilities. When you hear most folks talk about this or that being "possible," they are doing little more than expressing their own uncertainty or ignorance (e.g., "*I don't know! It's possible*."). This is not how a logician sees it. For us, the claim that "X is possible" is a claim of knowledge.

We know. Or rather, we will know for certain that something is possible—because we took the time to check. That's what the lower left part of a truth table provides: *a record of every possible* combination of truth values.

For an atomic statement, the total number of possible truth values is merely two: true or false. So, this means that if we wish to evaluate a compound statement that is made up entirely of just that one statement, there will also be only two possible combinations of truth values. We can start here, since this will be easy to do while illustrating many key features of how truth tables are set up.

So let's evaluate the following statement:

~ G

We start by setting up our truth table, first by laying out our four quadrants and placing the atomic statement(s) in the upper left corner.

G	

Now we will put the statement(s) we wish to evaluate in the upper right corner.

G	~ G

Now we list all possible truth values for the atomic(s) in our table and number these possibilities:

G	~ G
1. T	
2. F	

The statement we want to evaluate is made up of this atomic, so we can transfer over the possible truth values we listed on the left over to the right side, where we see the same atomic.

Like this:

G	~ G
1. T	T
2. F	F

I know this sounds a bit redundant right now, but that is only because this table is exceedingly simple. When we get to larger statements and larger truth tables, we will be rewarded for taking time on this very simple step.

Now we can consider the impact that our logical operator has on the truth of the overall statement. Negations are easy—they simply invert the truth value of whatever statement they deny. When G is true, denying that claim will be false. When G is false, denying that claim will be true. So, we put these inverted values *directly under* the ~ symbol. Like this:

G	~ G
1. T	F T
2. F	F T

This is a bit messy, so we will find it very helpful to clearly indicate which (if any) logical operator is the main connective of the statements we want to evaluate. Remember what we said about the main connective: one of its important qualities is that it is the bearer of truth for the statement. So when we indicate its importance on the truth table, we are graphically *showing the truth value for that compound* statement. We can either draw a small arrow over the main connective, draw a clear box around it, or both.

	G	~ G
1.	T	F T
2.	F	T F

I prefer to use a box without the arrow, but as we go further you may find that different ways of marking your table work especially well for you. You and/or your professor will come up with an agreeable method.

Now our truth table is complete. We have shown every possible relevant scenario in which the truth value of ~ G can be established. We call these scenarios "**cases**" or "**possible worlds**," and numbering them will help us to refer to them later, when we must justify our evaluations.

The completed table is nothing more than a picture of possible truth values. This can be used to answer questions about the statement we wanted to evaluate. We'll discuss this later, but for now just note that we need to first learn how to construct truth tables. Their importance will be discussed once we know how to set them up.

So far this has been pretty easy. However, if we have a compound statement to evaluate that has another atomic, then we must add that atomic statement to the upper left side. For example:

A v ~ G

This doesn't increase the number of possible truth values for the added statement; however, the number of *combinations* of truth values has increased.

For example, what are the possible combinations of truth values for A and B? After a bit of thought, most folks will be able to figure out that there are four possible combinations:

A is true, and G is true

A is false, and G is false

A is true, but G is false

A is false, but G is true

If you figured that out, good for you. Now what happens if we add a third atomic statement? For example:

M → (A v ~ G)

Now we have to come up with every possible combination of truth values for those three atomic statements.

Q: How many possible combinations are there?

Q: What are they?

Your job is to answer both questions. Do it correctly, without fail, and do it quickly.

Got it already?

How about now?

No?

Well, the answer to our first question is 8 combinations. Now, quickly tell me what those 8 combinations are and don't mess it up. Also, don't repeat yourself…

Go!

No?

Okay, you're going to have to get good at doing this…especially when we add *a fourth* atomic statement. Like this:

(H v M) → (A v ~ G)

Remember what you have to do. Answer the following questions:

Q: How many possible combinations are there?

Q: What are they?

Go!

Maybe you figured this out, but the answer to our first question with four atomics is 16 possible combinations. And yes, you have to tell me what they are and put them in your truth table (lower left quadrant). Sounds like fun.

If you don't think this is fun, you just haven't learned the method yet. We have a simple way to do this—really simple, as in a mechanical way to generate the answers that requires hardly any thought at all. How about that for a combination: *guaranteed success + little mental effort*

Ready? Here's how we do it.

FIRST STEP: Count the number of atomics. You will double the number of possible combinations every time you add an atomic statement.

Start with one atomic: you have two possibilities (T/F).

Add an atomic (now 2 atomics total) and double the possibilities: from 2 possibilities to 4 possibilities.

Add an atomic (now 3 atomics total) and double the possibilities: from 4 possibilities to 8 possibilities.

So on and so forth. Remember this phrase: "***double it up***"

That tells you how many cases (or possible worlds) you must account for in your table. Now that we know, we need to detail those cases and list them in the lower left quadrant of our table. This turns out to be surprisingly easy.

SECOND STEP: Find the atomic snuggled up against our vertical line. You will simply alternate T and F under it for as many cases as the table requires. Like this:

	M	A	G	M → (A v ~ G)
1			T	
2			F	
3			T	
4			F	
5			T	
6			F	
7			T	
8			F	

That's not hard. Now slide over one atomic and "*double it up*": You just alternated T and F every other line, so alternate every other *two* lines. Two true, two false, two true, two false…until all cases are completed. Like this:

	M	A	G	M → (A v ~ G)
1		T	T	
2		T	F	
3		F	T	
4		F	F	
5		T	T	
6		T	F	
7		F	T	
8		F	F	

Now slide over one atomic, and "double it up": You just alternated T and F every two lines, so alternate ever *four* lines. Four true, four false, four true, four false…until all cases are completed. Like this:

	M	A	G	M → (A v ~ G)
1	T	T	T	
2	T	T	F	
3	T	F	T	
4	T	F	F	
5	F	T	T	
6	F	T	F	
7	F	F	T	
8	F	F	F	

Since we only have 8 cases in this table, this is done in a single pair of *four* true / *four* false. However, if we had a table with 16 cases, we would see this pattern repeat more. We would also slide over to our last atomic, and once again "double it up" (alternating T/F every 8 lines). The method chugs along until we are done with all of our atomics. All we have to do is remember three rules:

Rule 1: The number of atomics is used to determine the number of cases (possible worlds)

Rule 2: Start with the atomic closest to the vertical line

Rule 3: "Double it up"

The rest of the table requires a bit more knowledge, but the initial set up is pretty mechanical. I should note that Rule 2 is just a convention. There is nothing logically necessary about starting with the atomic next to the vertical line. However, there is great advantage to holding this as a conventional rule for all of us to follow. This makes it easy for us to communicate the information on our table to others who are working on the same problem.

Communicating information is critical to justifying your evaluations. So, when you are asked to evaluate a statement or an argument *and justify* your findings, you will refer to a case (or cases) to do so. If you are communicating in your own "unique" way, the chances are good that you will not get credit for your answer because everyone else is looking at a different line than the one you claim supports your answer.

When we all set up our tables with the same conventions, we are all able to communicate clearly and precisely. Put differently, there are no unique snowflakes in logic. We all adhere to the same principles. So, the faster we get on board with the same methods, the easier it is to check that we are following those principles.

THIRD STEP: Now that we have identified and exhausted every possible combination of truth values for the atomic statements (that make up the statement[s] we want to evaluate), we can just transfer those values over to the lower right-hand quadrant. Like this:

	M	A	G	M	→	(A	v	~	G)
1	T	T	T	T		T			T
2	T	T	F	T		T			F
3	T	F	T	T		F			T
4	T	F	F	T		F			F
5	F	T	T	F		T			T
6	F	T	F	F		T			F
7	F	F	T	F		F			T
8	F	F	F	F		F			F

This might sound like a lot of redundant work, but this will make it very easy to complete the table. Before we can do that, we need to know *how our logical operators impact the truth values* of the component statements. We just saw this with negation, so let's review that and look at the rest of our logical operators.

Negation

Negation is simply a denial. So as we saw, you simply invert the value of whatever is negated.

	□	~	□
1	T	F	T
2	F	T	F

A Note on Annotation

In the above image, we are using our familiar □ (and later △ for the other operators) to indicate that a negation may apply to any statement at all, be it an atomic or a compound. However, we also used the □ (and later ∆ for the other operators) in the upper left-hand quadrant of the truth table. While □ and ∆ are our preferred statement *variables*, the reader should note that in this quadrant they can only ever be atomic statements. We never put compound statements in the upper left-hand quadrant of a truth table. In this quadrant, □ and ∆ can only represent atomic statements, whatever they are for the statements we wish to evaluate.

Conjunction

The conjunction is a union of two assertions. So think of this as a promise: I promise that we will do this **and** we will do that. To keep my promise (i.e., to ensure I speak the truth), I must fulfill both components. In other words, both sides need to be true in order for such a promise to be kept.

	□	△	□	•	△
1	T	T	T	T	T
2	T	F	T	F	F
3	F	T	F	F	T
4	F	F	F	F	F

Note that conjunctions are typically false. It really is hard to keep such a promise.

Memorizing the basic conditions of truth for conjunctions is easy—both sides have to be true for the conjunction to be true (any false conjunct makes the whole thing false).

Disjunction

The disjunction is a weak expression of possible options. Go back to the notion of keeping a promise. If I promise that at the very least, we will do this **or** we will do that, well heck, *it's kind of easy to keep that promise*. I can claim success in any of three scenarios: (1) we did this, (2) we did that, (3) we did both.

	□	△	□	v	△
1	T	T	T	T	T
2	T	F	T	T	F
3	F	T	F	T	T
4	F	F	F	F	F

Note that disjunctions are typically true. TIP: If you're going to make a promise, make it an Or-promise.

Memorizing the basic conditions of truth for disjunctions is easy—both sides have to be false for the disjunction to be true.

Material Conditional

The material conditional is a special relationship. We could (and do) casually call it a conditional, but the truth is that this is a logically specific relationship that is often quite different than what most people think of when they think "conditional." Thus, if you struggle with this, keep calling it a "material conditional" so that you remember it is a technical term and a special kind of conditional.

A material conditional is not a claim of causality, nor does it entail any special temporal elements (like "this happens before that"). Many English conditional expressions do include these extra senses of the words, but this is not translated over to the logical operator we know as the material conditional. What our conditional does is nicely capture the relationship of necessary or sufficient conditions. So, let's consider a story:

> My dad used to make promises that were dependent on me doing my chores. I was the family dishwasher. I also liked to go rent movies (when there were movie rental stores and we didn't have streaming services). So, dear old Dad would often promise, "If you do the dishes, we'll go to rent movies." *Grand idea Dad! Put the ball in my court*.
>
> On some days, I did the dishes. On other days, I dropped the ball and didn't do dishes. So what do you think dear old Dad did? Well, on some days, Dad took me to get the movies, and on other days he did not.

Q: On which days can we accuse Dad of speaking falsely?

Let's consider **only the days I did** the dishes.

Day 1: He takes me to get movies. *Yes!* Dad spoke the truth!

Day 2: He didn't take me to get movies. *What the!?!* Dad lied to me…I can't believe it, but he spoke falsely.

Now let's consider **only the days I did not** do the dishes.

Day 3: He takes me to get the movies anyway. Can I really be mad at Dad? *Nah*. (more on this in a bit)

Day 4: He didn't take me to get movies. *Well, what was I expecting?* I mean, after all, I didn't do the dishes. Can I accuse Dad of lying? No. So I'll accept that Dad spoke the truth.

If you had to do the dishes in your house, then the material conditional starts to make sense. The trouble for some is the Day 3 scenario. I did NOT do the dishes, so what am I to make of Dad taking me to get the movies anyway? I feel like he misled me. But did he? What was it that Dad said?

If you do the dishes, then we'll go rent movies.

From our work on translations, we see that "doing the dishes" is in the antecedent position (the □). This means that doing the dishes is in the place of the *sufficient condition*—it is supposed to be what logically *guarantees* that the △ claim is true. Also from our work on translations, we can see that this means that doing the dishes was not put in the place of the necessary condition (that's the △ part, and in this case it is the claim about going to get the movies). Now we can go back to Dad.

Q: Did dear old Dad ever say doing the dishes was *necessary*?

No, he did not. He said *it was sufficient* (he never said it was necessary). So, when he takes me to get the movies even though I did not do the dishes, he did not propose a false promise. He never said that doing the dishes was necessary for getting the movies; he merely assured me it was sufficient to get him off the couch to drive over there. I may have heard him incorrectly—that's always a strong possibility—but I cannot accuse *him* of misleading me simply because *I failed* to listen carefully. That shortcoming is on me.

Thus, the basic conditions of truth for the material conditional are as follows:

	□	△	□	→	△
1	T	T	T	T	T
2	T	F	T	F	F
3	F	T	F	T	T
4	F	F	F	T	F

The easy way to memorize the basic conditions of truth for the material conditional is to remember that you can really only pin a falsehood on it when the antecedent, □, is satisfied (i.e., is true) but the consequent △ was not delivered (i.e., is false).

Material Biconditional

Remember that biconditionals are what many people hear when someone says a conditional statement. They hear a very tight relationship between both clauses (the □ and the △). Our story of doing the dishes illustrates this well. We may *have heard* Dad make a biconditional promise. We thought he was saying:

I'll take you to get the movies *if* **and** *only if* you do the dishes.

That sounds like a Dad-promise. If he *had* made that promise, then we would expect the truth of his promise to look different than what we saw with the material conditional. And so it is.

	□	△	□	↔	△
1	T	T	T	T	T
2	T	F	T	F	F
3	F	T	F	F	T
4	F	F	F	T	F

Now we see in Case 3 (a.k.a. "Day 3" in our story) that there really is something shifty about Dad's promise if he was planning on going to get the movies all along. In making this biconditional promise, he spoke falsely! Sure, I may not mind (I got the movies), but I sure can't trust what that guy says. He said doing the dishes was both sufficient and necessary. *How can that be true!* How can it also be a *necessary* condition if he gets the movies anyway? Answer: it can't; Dad speaks with a forked tongue.

The easy way to memorize the conditions of truth for a biconditional is to remember:

Biconditionals are true when both sides *share the same* truth value, false if there is a *difference* on each side.

Overview of the Basic Conditions of Truth for the Logical Operators

We can capture all of this information in a single truth table for easy reference. You need to have this in your back pocket. We'll use "A" and "B" as placeholders for □ and △ to keep our table small. This is just a quick reference table that you must memorize:

	□	△	~	□	,	□	•	△	,	□	∨	△	,	□	→	△	,	□	↔	△
1	T	T	F	T		T	T	T		T	T	T		T	T	T		T	T	T
2	T	F	F	T		T	F	F		T	T	F		T	F	F		T	F	F
3	F	T	T	F		F	F	T		F	T	T		F	T	T		F	F	T
4	F	F	T	F		F	F	F		F	F	F		F	T	F		F	T	F

Calculating Truth Values for Statements

Have you memorized the truth table for the basic conditions of truth for all of the logical operators? If not, read no further. Stop what you are doing. Go back and reread. You must memorize the truth table for the basic conditions of truth of the logical operators. If you haven't done so, then from here on out, you are really wasting your time.

I know that culturally speaking, many people look down on rote memorization as an approach to learning that is somehow stultifying and soul-crushing. And sure, rote memorization can be soul-crushing if it's over-emphasized. But sometimes there is just no other way. Sometimes, as with much of the work in doing truth tables, it is actually *soul-strengthening*. Memorizing information requires you to focus. It can also feel good to flex your memory and see what it can do. If you don't memorize the truth tables, you will likely find logic to be confusing and depressing. If you do memorize them correctly, you will find much of what follows to be really easy.

One hugely important aspect of sentential logic is the calculation of truth values. Once you've memorized the truth table definitions for the operators, this turns out to be incredibly easy to do. It can also be somewhat relaxing. Just think of yourself as a mindless computer that takes certain inputs and generates certain outputs.

Consider the following statement:

(P → Q) ↔ (~ R → Q)

Earlier we saw that sentential logic is truth functional, which just means that the truth value of any complex statement (such as this one) is determined by the truth values of the simple statements it contains. The basic idea of truth functionality is simple: If you know what truth values to attach to the atomics in the statement above, then you can easily figure out what the truth value of the whole statement must be.

In order to see how this works, let's just stipulate:

P: true

Q: true

R: false

What would the truth value of the whole statement be? The statement as a whole is a biconditional, since the main operator is a triple bar.

(P → Q) ↔ (~ R → Q)

If we knew whether the chunk on each side were true or false, then we could consult the truth table for the triple bar to see what the the truth value of the whole statement must be. (And that is why it's so important to memorize the truth table definitions for the logical operators.)

So, let's take each chunk on its own. The chunk on the left is:

P → Q

We've stipulated that "P" and "Q" are both true, so "P → Q" must be true. So the left-hand chunk is true.

(P → Q) ↔ (~ R → Q)
 T

The right-hand chunk is just a bit more complicated. We know that "Q" is true. We've also stipulated that "R" is false. But that means that "~ R" is true. Since both "~ R" and "Q" are true, it follows that "~ R → Q" is also true. So the right-hand chunk is true as well.

(P → Q) ↔ (~ R → Q)
T T

Now recall that a biconditional statement basically asserts that the chunk on the left and the chunk on the right have the same truth value. Here that is indeed the case: the left-hand chunk and the right-hand chunk are both true. So the whole statement is true. We can indicate this by placing a "T" under the main operator.

(P → Q) ↔ (~ R → Q)
T T T

One reason why it's important to practice these truth value calculations is that this is the key to filling in truth tables for propositions. Suppose we wanted to do a full-blown truth table for the above statement. It would start out looking like this:

	P	Q	R	(P	→	Q)	↔	(~	R	→	Q)
1	T	T	T	T		T			T		T
2	T	T	F	T		T			F		T
3	T	F	T	T		F			T		F
4	T	F	F	T		F			F		F
5	F	T	T	F		T			T		T
6	F	T	F	F		T			F		T
7	F	F	T	F		F			T		F
8	F	F	F	F		F			F		F

Here we have already filled out the lower left quadrant with all possible combinations of truth values for the three atomics. We have also transferred those values over to the lower right quadrant for easy reference. Now we need to work through the logical structure of the statement as just discussed.

Working through Logical Structure

The first thing we need to do is mark the main connective and/or its eventual column. This helps us recognize the main components of the statement. Without knowledge of the main connective, you are lost. You have zero chance of successfully completing the table. So best to review that concept.

Practice your statement trees until you are very, very good at recognizing the logical structure of a statement. When you do a truth table, you are effectively *working your way up* through a statement tree: starting with the truth values for the atomics, then moving up through the statement, until you finally get to the very last operator to create the overall statement.

We start on our table by transferring those truth values to the atomics. This has been done above. So next step is to move on to the logical operators who join together or apply to *nothing more than atomics*. We can start with the left side of the biconditional.

The left side is a conditional statement. So the truth of this side will conform to the general conditions of truth for material conditional statements (which, of course, you have memorized by now). We indicate the truth value of this little conditional by putting its truth value directly under the arrow. Like this:

	P	Q	R	(P	→	Q)	↔	(~	R	→	Q)
1	T	T	T	T	T	T			T		T
2	T	T	F	T	T	T			F		T
3	T	F	T	T	F	F			T		F
4	T	F	F	T	F	F			F		F
5	F	T	T	F	T	T			T		T
6	F	T	F	F	T	T			F		T
7	F	F	T	F	T	F			T		F
8	F	F	F	F	T	F			F		F

The next statement we would like to do is the right side of the biconditional. This too is a little conditional statement. However, this arrow joins an atomic and a negation. So, we cannot fill in that arrow until we know the truth value of the antecedent. Put differently, the right side has the following form:

□ → Δ

We will know the truth of that condition when we know the truth values of □ and Δ. The Δ is an atomic, but the □ is a compound statement. So we cannot fill in the → column until we have a proper column for □.

In this case □ is a negation (specifically, inside the box is the statement "~ R"). Fortunately, this tilde applies *directly to an atomic* statement. So we can fill in the column for the ~.

	P	Q	R	(P	→	Q)	↔	(~	R	→	Q)
1	T	T	T	T	T	T		F	T		T
2	T	T	F	T	T	T		T	F		T
3	T	F	T	T	F	F		F	T		F
4	T	F	F	T	F	F		T	F		F
5	F	T	T	F	T	T		F	T		T
6	F	T	F	F	T	T		T	F		T
7	F	F	T	F	T	F		F	T		F
8	F	F	F	F	T	F		T	F		F

Now we have the right side of the biconditional ready to complete. Our column *under the tilde* bears the truth of □, so we use *that column* along with the column under the "Q" to determine the truth values in the column under the right side →.

	P	Q	R	(P	→	Q)	↔	(~	R	→	Q)
1	T	T	T	T	T	T		F	T	T	T
2	T	T	F	T	T	T		T	F	T	T
3	T	F	T	T	F	F		F	T	T	F
4	T	F	F	T	F	F		T	F	F	F
5	F	T	T	F	T	T		F	T	T	T
6	F	T	F	F	T	T		T	F	T	T
7	F	F	T	F	T	F		F	T	T	F
8	F	F	F	F	T	F		T	F	F	F

Our table is almost done. We can now see the column of truth values under each side of our biconditional. *These two columns* will be used to determine the column of truth values under the ↔.

	P	Q	R	(P	→	Q)	↔	(~	R	→	Q)
1	T	T	T	T	T	T		F	T	T	T
2	T	T	F	T	T	T		T	F	T	T
3	T	F	T	T	F	F		F	T	T	F
4	T	F	F	T	F	F		T	F	F	F
5	F	T	T	F	T	T		F	T	T	T
6	F	T	F	F	T	T		T	F	T	T
7	F	F	T	F	T	F		F	T	T	F
8	F	F	F	F	T	F		T	F	F	F

Of course, we have memorized the basic conditions of truth for the ↔, so this should be easy.

	P	Q	R	(P	→	Q)	↔	(~	R	→	Q)
1	T	T	T	T	T	T	T	F	T	T	T
2	T	T	F	T	T	T	T	T	F	T	T
3	T	F	T	T	F	F	F	F	T	T	F
4	T	F	F	T	F	F	T	T	F	F	F
5	F	T	T	F	T	T	T	F	T	T	T
6	F	T	F	F	T	T	T	T	F	T	T
7	F	F	T	F	T	F	T	F	T	T	F
8	F	F	F	F	T	F	F	T	F	F	F

Now we have a complete picture of every possible case in which this statement is true and when it is false. The obvious question is: *What are we going to do with this picture?*

Using Truth Tables to Evaluate Logical Properties

Recall in Chapter 1 that we learned about important logical properties of statements. The savvy student remembers that all of these were defined in terms of possibility. So, naturally a truth table is an ideal tool to use for these properties.

Evaluating Individual Statements

The truth table we just built demonstrates every possible outcome for our previous sample statement:

(P → Q) ↔ (~ R → Q)

Now we can ask questions of possibility. The first, and easiest, might be:

Q: Is this statement a logical tautology, contradiction, or contingent statement?

If you don't remember what these properties are, here's the cheat sheet version:

Tautology: a statement which is true under all conditions

Contradiction: a statement which is false under all conditions

Contingent statement: a statement which is true under some conditions and false under other conditions

A truth table provides us with "the conditions" that are important. We call these the "cases" or "possible worlds." So each numbered case represents a specific condition of a possible world. When we look at a truth table, we are looking at the statement's main connective column to answer this question.

Q: What do you see in the MC column? Is it all Ts, all Fs, or is it a combination?

	P	Q	R	(P	→	Q)	↔	(~	R	→	Q)
1	T	T	T	T	T	T	T	F	T	T	T
2	T	T	F	T	T	T	T	T	F	T	T
3	T	F	T	T	F	F	F	F	T	T	F
4	T	F	F	T	F	F	T	T	F	F	F
5	F	T	T	F	T	T	T	F	T	T	T
6	F	T	F	F	T	T	T	T	F	T	T
7	F	F	T	F	T	F	T	F	T	T	F
8	F	F	F	F	T	F	F	T	F	F	F

We see that there is a combination of truth values. We don't see a straight column of nothing but Ts (nor a column of nothing but Fs). If we did see a pure column of Ts, we would say this statement cannot be false; it is always true under all conditions (i.e., it's a tautology or logically true statement). We would say the opposite if it were a column of pure Fs (i.e., it's a contradiction or logically false statement). However, in this case we do not see those kinds of columns. We see a column in which sometimes the statement is true and sometimes it is false. So, we say that the statement is a contingent (or logically indeterminate) statement.

We might even go deeper into this table. If a friend doubted that this statement is ever true, we can provide them with a very specific set of conditions in which the statement would indeed be true. We can point to specific cases in which it is true. This can give us some sense of what the world must look like in order for the statement to be true or false.

Evaluating Sets of Statements

Many times, we want to know something about a group of statements. They may not even be an argument, or they may be small parts of an argument. No matter, we just want to know things about a given group. Truth tables can help us explore the logical properties of these groups.

Logical Equivalence

For example, consider the following:

> Your friend makes a claim, then later says something different, but insists they said the same thing. In the morning your friend said, "*Sure. If I go to the store, I'll get some milk*." In the afternoon, your friend said, "*Okay already! Either I don't go to the store, or I'll get some milk.*" You think your friend is changing their story to suit themselves. But your friend may defend themselves by insisting that even though they said it differently, their second statement really meant "the same" thing. Well, does it?

A truth table can discover if two statements are logically equivalent. Again, if Chapter 1 content does not spring to mind immediately, here's the cheat sheet version:

> **Logical Equivalence**: two statements are logically equivalent if and only if they always have the same truth value as one another under all conditions

Since a truth table demonstrates all conditions, let's go check the table.

We need to first translate our friend's statement. This will do:

> If I go to the store, I'll get some milk.S → M
>
> Either I don't go to the store, or I'll get some milk. ~ S v M

We can put multiple statements in a truth table's upper right-hand quadrant by simply distinguishing them with commas. Like this:

	S	M	S	→	M	,	~	S	v	M
1	T	T								
2	T	F								
3	F	T								
4	F	F								

With only two atomics, this is a small truth table. The finished version looks like this:

	S	M	S	→	M	,	~	S	v	M
1	T	T	T	T	T		F	T	T	T
2	T	F	T	F	F		F	T	F	F
3	F	T	F	T	T		T	F	T	T
4	F	F	F	T	F		T	F	T	F

With a complete table we can look at the two columns under the main connective of each statement.

Q: Are they *identical columns*, or is there *at least one variation* in the truth values?

Now we can see that our friend really did say the same thing, insofar as their statements reflect the same account of the world. These two statements will always have the same truth value under any and all conditions.

CAUTION: Truth tables can be used this way only if we put both statements in the same table. You cannot reliably make one table for the first statement and another table for the second statement. This would run the risk of demonstrating very different possible combinations of truth values (i.e., different possible worlds). If this happens, we cannot make accurate claims about what is and is not possible. We must have all statements under evaluation *in a single truth table* to make those judgments.

Logical Consistency and Inconsistency

Let's look at our other concepts for groups of statements. Remember logical consistency and inconsistency? No worries, here's the cheat sheet version:

> **Logical Consistency**: a set of statements is logically consistent if and only if it is possible for all members of the set to be true under the same conditions
>
> **Logical Inconsistency**: a set of statements is logically inconsistent if and only if it is not possible for all members of the set to be true under the same conditions

Truth tables make short work of any question regarding these concepts. Let's look at an example.

Consider the following statements:

~ K → W

~ (W v K)

W · K

Q: Are these statements logically consistent or are they inconsistent? When we put them all together in a truth table, we get the following:

	K	W	~	K	→	W	,	~	(W	v	K)	,	W	•	K
1	T	T	F	T	T	T		F	T	T	T		T	T	T
2	T	F	F	T	T	F		F	F	T	T		F	F	T
3	F	T	T	F	T	T		F	T	T	F		T	F	F
4	F	F	T	F	F	F		T	F	F	F		F	F	F

Our question is in regard to consistency, so we need to look at the columns under the main connective for all three statements. We look across the table, case by case, at each truth value under the main connectives. We are looking to see if there is any case (at least one) in which all three statements are true. This would demonstrate that it is possible for them to be true under the same conditions.

In this case, we do not see a single case in which all statements are true. That is, we never see the possibility of all statements being true, i.e., it is impossible for all members of this set of statements to be true under the same conditions. The set is inconsistent.

Let's try another example. Consider the following statements:

~ B v E

~ (E v ~ M) → ~ B

M · ~ E

Q: Are these statements logically consistent or are they inconsistent? When we put them all together in a truth table, we get the following:

	B	E	M	~	B	v	E	,	~	(E	v	~	M)	→	~	B	,	M	•	~	E
1	T	T	T	F	T	T	T		F	T	T	F	T	T	F	T		T	F	F	T
2	T	T	F	F	T	T	T		F	T	T	T	F	T	F	T		F	F	F	T
3	T	F	T	F	T	F	F		T	F	F	F	T	F	F	T		T	T	T	F
4	T	F	F	F	T	F	F		F	F	T	T	F	T	F	T		F	F	T	F
5	F	T	T	T	F	T	T		F	T	T	F	T	T	T	F		T	F	F	T
6	F	T	F	T	F	T	T		F	T	T	T	F	T	T	F		F	F	F	T
7	F	F	T	T	F	T	F		T	F	F	F	T	T	T	F		T	T	T	F
8	F	F	F	T	F	T	F		F	F	T	T	F	T	T	F		F	F	T	F

Again, we look across the table to read for consistency, focusing only on the main connectives of the three statements in our set. Going case by case, we see that it is rarely the case that all three statements in the set are true under the same conditions. This almost never happens. Rarely, but *not impossible*!

Indeed, in case #7 we see that *it is possible* for all three statements in our set to be true under the same conditions. What conditions? The condition of the world in which B and E are false while M is true.

	B	E	M	~	B	v	E	,	~	(E	v	~	M)	→	~	B	,	M	•	~	E
1	T	T	T	F	T	T	T		F	T	T	F	T	T	F	T		T	F	F	T
2	T	T	F	F	T	T	T		F	T	T	T	F	T	F	T		F	F	F	T
3	T	F	T	F	T	F	F		T	F	F	F	T	F	F	T		T	T	T	F
4	T	F	F	F	T	F	F		F	F	T	T	F	T	F	T		F	F	T	F
5	F	T	T	T	F	T	T		F	T	T	F	T	T	T	F		T	F	F	T
6	F	T	F	T	F	T	T		F	T	T	T	F	T	T	F		F	F	F	T
7	F	F	T	T	F	T	F		T	F	F	F	T	T	T	F		T	T	T	F
8	F	F	F	T	F	T	F		F	F	T	T	F	T	T	F		F	F	T	F

So this set is logically consistent. Our justification would be "case #7" (or any other case in which this condition was met, if it were met by multiple cases). Note that since consistency merely requires "*the* possibility" of all true statements, a correct and complete justification does not require that we enumerate *every* case in which it occurs. Citing just a single case is sufficient.

Take note of an important principle of reading truth tables. Completed truth tables demonstrate possibility and impossibility differently:

Truth tables demonstrate "POSSIBLE" with *at least one case* that you can point to with the quality under question.

Truth tables demonstrate "IMPOSSIBLE" with the absence of the quality in the *entire* table.

Formal Validity and Invalidity

Of course, by far the most important property that interests us occurs when the set of statements is structured as an argument. We can put arguments in a truth table by modifying how we put regular sets in a table. We'll use commas to distinguish the premises from one another and then a single backslash to separate the conclusion from the other statements. Consider the following argument:

1. (L v B) → H
2. S ↔ L
3. ~ B · H

∴ ~ S

We can set this up in a single truth table to determine if the argument is valid or invalid. Our cheat sheet definitions will help us focus on what to look for in the completed table.

Validity: an argument is valid if and only if it is **not possible** for all the premises to be true and the conclusion false under the same conditions

Invalidity: an argument is invalid if and only if it **is possible** for all the premises to be true and the conclusion false under the same conditions

Here is our completed truth table for the argument:

	L	B	H	S	(L	v	B)	→	H	,	S	↔	L	,	~	B	•	H	/ ∴	~	S
1	T	T	T	T	T	T	T	T	T		T	T	T		F	T	F	T		F	T
2	T	T	T	F	T	T	T	T	T		F	F	T		F	T	F	T		T	F
3	T	T	F	T	T	T	T	F	F		T	T	T		F	T	F	F		F	T
4	T	T	F	F	T	T	T	F	F		F	F	T		F	T	F	F		T	F
5	T	F	T	T	T	T	F	T	T		T	T	T		T	F	T	T		F	T
6	T	F	T	F	T	T	F	T	T		F	F	T		T	F	T	T		T	F
7	T	F	F	T	T	T	F	F	F		T	T	T		T	F	F	F		F	T
8	T	F	F	F	T	T	F	F	F		F	F	T		T	F	F	F		T	F
9	F	T	T	T	F	T	T	T	T		T	F	F		F	T	F	T		F	T
10	F	T	T	F	F	T	T	T	T		F	T	F		F	T	F	T		T	F
11	F	T	F	T	F	T	T	F	F		T	F	F		F	T	F	F		F	T
12	F	T	F	F	F	T	T	F	F		F	T	F		F	T	F	F		T	F
13	F	F	T	T	F	F	F	T	T		T	F	F		T	F	T	T		F	T
14	F	F	T	F	F	F	F	T	T		F	T	F		T	F	T	T		T	F
15	F	F	F	T	F	F	F	T	F		T	F	F		T	F	F	F		F	T
16	F	F	F	F	F	F	F	T	F		F	T	F		T	F	F	F		T	F

Remember, a truth table shows *impossibility* when the condition is never met in the *entire table*. A truth table shows *possibility* with the condition met in just a *single case*.

So what do you see in this table? Do you see at least one case in which all the premises are true (i.e., under the main connectives, you see a T) and the conclusion is false (there is an F value under its main connective)? If you do see that, then you know the argument is invalid. You should then cite the case number as the *justification for your judgment* that the argument is invalid. If you never see this, then you make that claim as your justification.

Look again at the table to find your answer.

This argument is INVALID, case #5. We can see this clearly when we circle the case or put emphasis on it, like this:

	L	B	H	S	(L	v	B)	→	H	,	S	↔	L	,	~	B	•	H	/∴	~	S
1	T	T	T	T	T	T	T	T	T		T	T	T		F	T	F	T		F	T
2	T	T	T	F	T	T	T	T	T		F	F	T		F	T	F	T		T	F
3	T	T	F	T	T	T	T	F	F		T	T	T		F	T	F	F		F	T
4	T	T	F	F	T	T	T	F	F		F	F	T		F	T	F	F		T	F
5	T	F	T	T	T	T	F	T	T		T	T	T		T	F	T	T		F	T
6	T	F	T	F	T	T	F	T	T		F	F	T		T	F	T	T		T	F
7	T	F	F	T	T	T	F	F	F		T	T	T		T	F	F	F		F	T
8	T	F	F	F	T	T	F	F	F		F	F	T		T	F	F	F		T	F
9	F	T	T	T	F	T	T	T	T		T	F	F		F	T	F	T		F	T
10	F	T	T	F	F	T	T	T	T		F	T	F		F	T	F	T		T	F
11	F	T	F	T	F	T	T	F	F		T	F	F		F	T	F	F		F	T
12	F	T	F	F	F	T	T	F	F		F	T	F		F	T	F	F		T	F
13	F	F	T	T	F	F	F	T	T		T	F	F		T	F	T	T		F	T
14	F	F	T	F	F	F	F	T	T		F	T	F		T	F	T	T		T	F
15	F	F	F	T	F	F	F	T	F		T	F	F		T	F	F	F		F	T
16	F	F	F	F	F	F	F	T	F		F	T	F		T	F	F	F		T	F

Note that if I had seen *more than one case*, I would not need to cite all cases in which the premises are all true and the conclusion false. I just need to cite one case to demonstrate that it is indeed possible for this to happen. Finding many cases would not make the argument "more invalid," so it would not strengthen my evaluation.

A Cautionary Tale

Many students new to truth tables struggle with the following observation:

I see a case in which all the statements in the argument are true!

That's great. However, what do you think this tells you about the argument? Try this quick quiz:

T/F: An argument that has all true statements is valid.

Go ahead and take a moment to think about your answer.

We'll wait.

Having all true statements in your argument sounds like a good thing. I mean, we are speaking the truth. Isn't that a good thing?

Well, the answer really is: *it depends*.

We have to keep in mind what our intention is with what we say. If *all I want* to do is accurately report on the conditions of the world, then yes, having a bunch of true statements is a very good thing. However, if I want to do more with what I say, if *I want* to use some statements as support for the truth of another, then having all true statements is not necessarily such a resounding success. If this is my intention, then I need to judge my efforts by a higher standard than simply having the truth.

Take another look at the previous truth table. Do you see any conditions in which we have *all true* statements in our argument?

	L	B	H	S	(L	v	B)	→	H	,	S	↔	L	,	~	B	•	H	/∴	~	S
1	T	T	T	T	T	T	T	T	T		T	T	T		F	T	F	T		F	T
2	T	T	T	F	T	T	T	T	T		F	F	T		F	T	F	T		T	F
3	T	T	F	T	T	T	T	F	F		T	T	T		F	T	F	F		F	T
4	T	T	F	F	T	T	T	F	F		F	F	T		F	T	F	F		T	F
5	T	F	T	T	T	T	F	T	T		T	T	T		T	F	T	T		F	T
6	T	F	T	F	T	T	F	T	T		F	F	T		T	F	T	T		T	F
7	T	F	F	T	T	T	F	F	F		T	T	T		T	F	F	F		F	T
8	T	F	F	F	T	T	F	F	F		F	F	T		T	F	F	F		T	F
9	F	T	T	T	F	T	T	T	T		T	F	F		F	T	F	T		F	T
10	F	T	T	F	F	T	T	T	T		F	T	F		F	T	F	T		T	F
11	F	T	F	T	F	T	T	F	F		T	F	F		F	T	F	F		F	T
12	F	T	F	F	F	T	T	F	F		F	T	F		F	T	F	F		T	F
13	F	F	T	T	F	F	F	T	T		T	F	F		T	F	T	T		F	T
14	F	F	T	F	F	F	F	T	T		F	T	F		T	F	T	T		T	F
15	F	F	F	T	F	F	F	T	F		T	F	F		T	F	F	F		F	T
16	F	F	F	F	F	F	F	T	F		F	T	F		T	F	F	F		T	F

With the emphasis here we can easily see that, yes, case #14 depicts a situation in which all the statements are true. Cool!

Q: What does that tell you about *the argument*?

Q: Does this tell us if the argument is *valid*?

Well, look again at the truth table. Notice that case #5 is still there! The table hasn't changed simply because we looked at case #14. The table shows that even if you are living in a world described by case #14 and all your statements are in fact true, it is *still possible* for all the premises to be true and the conclusion false. The possibilities for this argument are the same. Put differently, you might live in the world in which you have a bunch of true statements coming out of your mouth, but the possibility that case #5 describes for those statements isn't going anywhere. Your argument is still invalid.

Shortened Truth Tables

We have seen how a complete truth table can help us quickly evaluate statements, sets of statements, and arguments for their formal properties. Often enough, the tables are of manageable size to set up quickly and complete without too much burden. However, once we get past four atomics, truth tables start to get a bit unwieldy in size. Doing a 32- or 64-line or larger truth table is good for the soul—it builds mental toughness. However, as a practical matter, it is very time-consuming. At the end of it all, *we really end up only looking at those few lines that demonstrate the possibility of a combination of truth values* that interest us. So, this is a clue as to how we can shorten up truth tables.

A shortened truth table requires a very high level of mastery over the basic conditions of truth for our logical operators. The "short" in shortened truth table makes it sound easy, and thus appealing; however, you will likely make serious errors if you are not very well-versed in the basic conditions of truth. You are likely to be more successful completing a full truth table than a shortened one until you become really familiar with how the logical operators work. Consider this a warning.

We can use shortened truth tables for most any logical property we would like to investigate. However, the most common use is to test for validity. We remember that validity is defined as a certain "impossibility" which requires an entire truth table to demonstrate. However, invalidity is defined as a certain type of "possibility" which requires only one line of a truth table. That line is as follows:

P1	P2	P3	/	C
T	T	T		F

Whatever our premises and conclusion, we know that under the main connective of each must appear this specific combination of truth values to show that the argument is invalid. I don't really care what the case number is on a truth table; I just want to know if this is a possibility. So…let's just set them up that way, *and then* use our knowledge of the basic conditions of truth to see if we can make those values work out.

Consider our previous example:

1. (L v B) → H
2. S ↔ L
3. ~ B · H

∴ ~ S

If we lay this out as it would appear in the upper right-hand quadrant of a truth table, it looks like this:

(L v B) → H , S ↔ L , ~ B · H / ~ S

We can then put the possible combination of truth values that we want to test under each statement's main connective. Since we're testing for validity, it would look like this:

(L v B) → H , S ↔ L , ~ B · H / ~ S
T T T F

If this is possible, then we know the argument is invalid.

Our next step is to see if any truth values become "forced under this possibility."

Q: Can we look at any of these values and see that some other statement "must be" a specific truth value?

We should see one immediately drop into our lap. If our conclusion is ~ S and we are testing the possibility that it is a false statement, then S *must be* a true statement. So we fill this in everywhere we see an S appear:

(L v B) → H , S ↔ L , ~ B · H / ~ S
T T T T F T

Notice, because the conclusion forced a value for S, we need to keep that value for all S instances in the argument. So this is why we included this in our table for the second premise.

Now we see if this new value also forces any more values. Since our second premise is a biconditional that is true, we know both sides of the biconditional *must* share the same truth value. We now know that the left side is true, so this forces the right side to be true (in order to keep the T value under the double arrow). Like this:

(L v B) → H , S ↔ L , ~ B · H / ~ S
T T T T T T F T

As before, we need this same value for all instances of L. The same process repeats. We look for forced truth values with each set of values that we now know "must be the case" under our test for invalidity.

Sometimes we see multiple forced values. For example, the savvy student has likely already picked up on two more forced values. The third premise is an And statement that is true. So we know that each conjunct must be true. We also know that the first premise is a conditional statement whose antecedent is itself a disjunction (L v B) which has a true conjunct (the "L"). That forces the "v" to be true as well. Before you get all eager to start dropping these values in, you might want to slow down. Doing a shortened truth table is most successful when we take each small step in its turn. Often in logic, slow is faster. Slow is also more accurate. So, let's proceed slowly.

Remember that we just said our third premise, (~ B · H), is a true And statement, and this means both conjuncts must be true. Let's do just that one for now:

(L v B) → H , S ↔ L , ~ B · H / ~ S
T T T T T T T T T F T

Don't forget that the "H" appears twice in the premises, so once we know that it is true in the third premise we must use that value in all other appearances of the "H" in the table. So far, so good...

The savvy student has now started to see where this is going. *So far, so good*... That's the worry. If *we are able* to fill in all the truth values for every atomic statement *without any contradiction*, then we will be able to *go so far* as to say it is indeed possible for all the premises to be true and the conclusion false. We are trying to see if these values pan out, or if we run into a situation where we cannot fill in truth values without violating our principle of bivalence (i.e., no statement can be both true and false under the same conditions). Let's see if we can keep going.

The very savvy student has already sniffed out something quirky about our first premise. That premise is a conditional statement whose consequent is true. This means that it doesn't matter what the truth value of the antecedent is, since the conditional will be true all the same. We also know that we really only need to see if we can give a truth value to the atomic "B" statement that is in the antecedent.

Since the truth value of the antecedent no longer matters, we are pretty confident that we can give B any truth value that may be required if one is forced upon it somewhere else in the argument. So we look for other instances of the atomic B.

Of course, we find a B in the third premise. Moreover, the value for B there is forced. Because ~ B must be a true statement, the B must be false.

(L v B) → H , S ↔ L , ~ B · H / ~ S
T F T T T T T TF T T F T

Now we can see that the truth values of all the atomics have been found without creating any contradictions. We could fill in the truth value for (L v B) if we wanted, but as we realized earlier, it really doesn't matter. The little "v" is true, but the first premise was going to be true anyway. This means we have found *a possible scenario* in which all premises are true and the conclusion is false. Which scenario? The one in which the truth values of the atomics that define that scenario are as follows:

L B H S
T F T T

Under this case (the case # doesn't really matter), conditions are such that all premises are true and the conclusion is false. We have demonstrated that this is possible. Thus, the argument is invalid.

So what happens when the argument is valid? In this case, we will try as we just did to fill in all combinations of truth values. However, along the way we will realize that we cannot keep going so far. We get forced into claiming that the truth value for a given statement is both true *and* false. That cannot be possible, thus the argument is valid (because the only way to show all true premises and a false conclusion is to insist that the statements are not "really" those truth values).

Here's a simple example:

1. Y → G
2. Y

∴ G

Clearly, this is a valid argument; after all, it is nothing more than an instance of the valid form modus ponens (which we will use extensively in the next chapter). To prove that this is valid, we put it in a shortened truth table:

Y → G , Y / G
T T F

Since one of our premises is already an atomic, this forces all other instances of "Y" to share the same value. So we put that in:

Y → G , Y / G
T T T F

Of course, the same can be said of our conclusion, so that would have been a fine place to start as well. We chose the atomic Y, so let's stick with that for a bit.

When we put the truth value in for the Y that appears in the first premise, we should see that this forces a value for the G in that premise. After all, the first premise is a *true* conditional statement whose antecedent is true. This means the consequent *must be true* in order to maintain the T under the arrow. So we put that in:

Y → G , Y / G

T T T T F

So far, so good. Now we have a truth value for Y and we have a truth value for G. But wait! We have two truth values for the atomic G! In the premise, G must be true, but in the conclusion, it is F. We would have to change the truth value in the conclusion to a T, which is conventionally shown like this:

Y → G , Y / G

T T T T F/T

Now our shortened truth table demonstrates that *the only way* to make these values work out is for G to have two truth values at once. This *is not possible*. Thus, the argument is valid. Put differently, the only way to make a case in which all premises are true and the conclusion false requires something that is not possible.

In practice, when doing shortened truth tables, there may be several ways to show that a combination of truth values is not possible. So your friend's shortened truth table may appear different than yours, yet both tables are correct. The difference between the two will be the result of your each following up on a different forced value.

Last word on shortened truth tables: You do not have to do them.

If this all sounds like far too much thought to put into a truth table, fine. You can fill out a complete truth table without much effort; it just takes a bit of time. The full table will demonstrate the same possibility or impossibility. Full truth tables work, and they are not terribly hard to complete.

However, if you feel like you have the hang of the basic conditions of truth, shortened truth tables are kind of fun and more engaging than a full table. And yes, if we're talking about a 128-line truth table, shortening that kind of beast up will save a bit of time too.

7.

PROPOSITIONAL LOGIC

We have seen that with a crisp understanding of how to translate natural language expressions into formal symbolic expressions, we can use formal tools to evaluate many logical properties of statements. In this chapter, we will learn one of the most important methods for evaluating arguments, not simply for its accuracy but also for the many logic life lessons it imparts.

Natural Deduction

Truth tables provide a powerful tool for evaluating the validity of an argument. In one shot we capture all possibilities to help us evaluate the argument. However, this strength is also the source of its shortcoming.

A truth table works like a litmus test. It only shows you "after the fact" that something is or is not possible. In this case, the table reveals that "after the reasoning that led to the conclusion," one can say with confidence that the overarching inference between premises and conclusion can or cannot guarantee the preservation of truth.

While this is nice, this is also not terribly revealing of how we (or someone much smarter than us) *got to* the conclusion in the first place. Truth tables do not show this to anyone who might not immediately see it. An example may help illustrate the point.

Let's say I want to offer the following argument:

1 $(L \vee M) \cdot (L \vee S)$

2 $A \to {\sim}L$

3 $({\sim}S \to L) \to {\sim}M$

$\therefore {\sim}A$

Remember back in the first chapter we said that an inference is like "a leap of the mind." Logicians really only care if the leap is a secure move, one that preserves whatever truth we have in hand. Logicians don't care if the leap is large or small, only that it is secure. Yet some folks can make both large *and* secure leaps of mind. Everyone can jump, but only some can compete in the Olympics and make really large jumps. A *really* sharp mind may look at this and just "see" that the conclusion follows from the premises. Their mind jumps from those premises to the conclusion in a single bound.

"*Oh yeah, of course, that follows*."

Wow…that guy is sharp. Frankly, when I look at it, I do not just "get it" immediately. My mind can't jump that far. Sure, some people can do this; their minds can recognize in one shot that the conclusion really does follow from the premises. However, most folks cannot (so don't feel bad if you don't see it either). We're not all Olympic-level competitors in the inference jump.

What most of us can do is make tiny jumps. Those Olympic athletes can soar over 25 feet in one jump! But you know, if I want to get from here to some place 25 feet away, I can also do it. *I just walk.* Walking is a series of super-small jumps. For a split second you launch yourself, teetering over your foot on the brink of falling, only to elegantly land on the other foot. Nice. Even better, *I can go further* than 25 feet, and I don't even have to be a super athlete to do it. Natural deduction is like that.

While truth tables give us a picture of the argument for a litmus test of its validity, natural deduction provides us with a way to show something much closer to how we actually think (at least when we are being very careful and very precise). A method of natural deduction breaks down very large inferences into much smaller steps. Indeed, like walking, those steps are so small that they can seem rather easy: we might even say they are trivial, at least until we see how far they can take us.

Let's revisit our argument example. Let's say I wanted to help a friend out. My friend does not see how the conclusion follows—or at least they are skeptical that it follows. I need to demonstrate this to them (and no, I don't hang out with Olympic-level inference leapers). So, I need to take them through it *step* by *step*.

I can continue to number my statements but move my conclusion to the side to make this easier. I could first say something like this:

1	(L v M) • (L v S)	
2	A → ~ L	
3	(~ S → L) → ~ M	∴ ~ A
4	L v M	

With a puzzled look, my friend asks why I should believe that *this* claim is true. To help my friend, the exchange follows as such:

> **Me**: *Look, remember on line 1 when we said (L v M)* ***and*** *(L v S), so you see, that's why. Right? If both of those are true, the left side has to be true.*
>
> **My Friend**: *Oh, okay, I get that.*

Now I can continue like this:

1	(L v M) • (L v S)	
2	A → ~ L	
3	(~ S → L) → ~ M	∴ ~ A
4	L v M	
5	L v S	

Again, my friend wants to know why I would say that. So I offer:

Me: *Look, remember again on line 1 when we said (L v M)* ***and*** *(L v S). If both of those are true, the right side has to be true. So you see, that's why. Right?*

My Friend: *Oh right, right, I get that*.

This goes on a few times. I keep pointing my friend to the statements we previously agreed upon, and my friend works hard to figure out why we're on solid ground. Eventually, we get a full account. It might look like this:

1	$(L \lor M) \cdot (L \lor S)$	
2	$A \rightarrow \sim L$	
3	$(\sim S \rightarrow L) \rightarrow \sim M$	$\therefore \sim A$
4	$L \lor M$	
5	$L \lor S$	
6	$S \lor L$	
7	$\sim S \rightarrow L$	
8	$\sim M$	
9	L	
10	$\sim A$	

At each step along the way, my friend asks for a justification, and for each line I point out what my friend should look at to understand where my little inferences are coming from.

In a very *natural* way, I have taken a step-by-step approach to help my friend understand why the conclusion ultimately does follow from the truth of the premises. Now my friend is assured not just that the argument is valid, but that they can understand the line of reasoning that led to the conclusion. This is a major advantage over the simple litmus test of a truth table.

However, what we have seen thus far is not really a proof. In this little story, as I made my way to the conclusion, I made my friend do all of the heavy lifting. I asked him: *Don't you see? Don't you see? Right? Right?* My friend had to dig deep to figure out why each claim was ensured by the statements I was pointing out. They may have even needed to whip out a truth table to ensure these were solid moves. *That's not nice*. I made my friend do all the work!

Had I really offered a full proof, *I* would have done work of *justifying* each statement—not simply making my friend figure that part out for themselves.

To remedy this shortcoming, when we do a proof we will offer complete justification for each small step that leads us to a new statement. Each line of our proof will include a quick justification that my friend can easily follow. These justifications will invoke basic rules of logic.

These **"rules of logic"** are really just small **valid inference forms** that we know we can rely upon. And yes, you can use a truth table to verify that they express valid inferences.

Derivations, a.k.a. Symbolic Proofs

We will develop a system of natural deduction that takes advantage of our symbolic translations. We casually call these "logical derivations," "formal proofs," or "symbolic proofs." This method will make use of the syntax of our propositions, because we can directly apply the valid inference forms to them. Thus, this method is often referred to as "proofs in propositional logic."

A complete proof will contain:

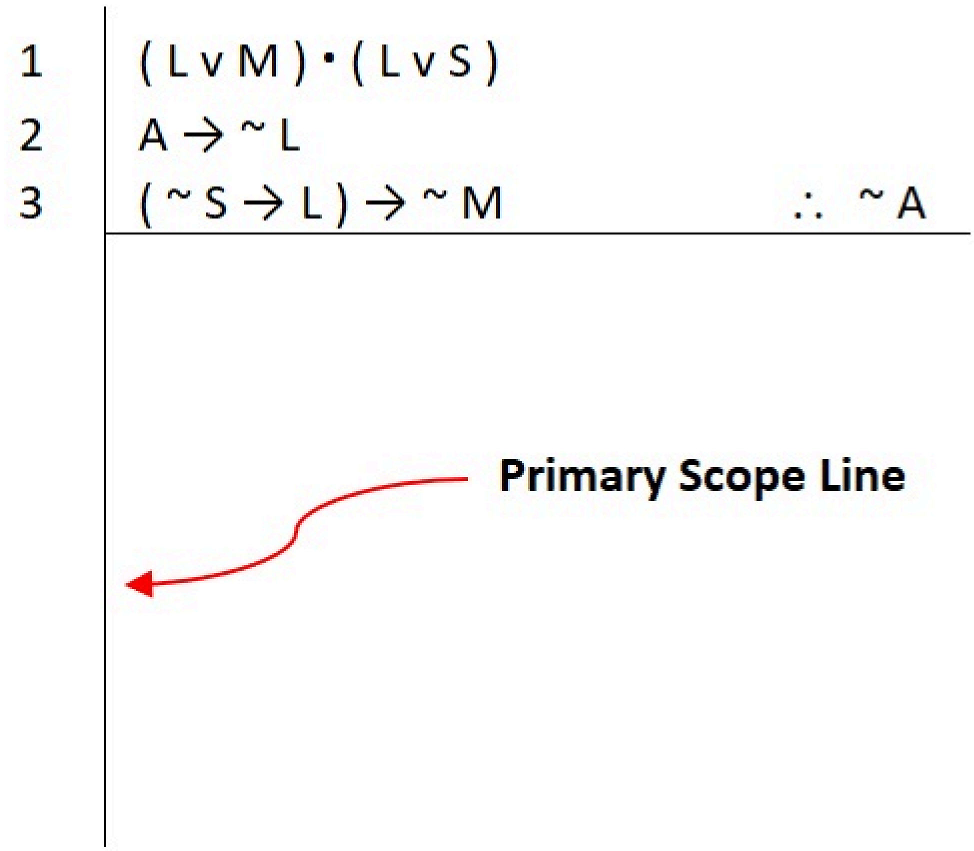
a. Our starting point: the premises we are given to treat as true statements
b. A notation of our final goal: the conclusion
c. Intermediate statements derived from our starting point: these are the small steps we took along the way
d. Justification for the truth of every statement made after the premises: these use the rules of logic and are abbreviated for ease of reference

Our method[1] will be organized in a specific and common way to make it easy to communicate our proof to other logicians. Here's our previous example with each line justified:

1. In this text, we will be using a modified "Fitch method" of annotation, named after logician Frederic Fitch, who developed this technique. Most contemporary introductory logic texts make use of some form of Fitch method in which the "Fitch bar" (the horizontal line separating premises from derived statements), vertical scope lines, and indentation of statements and justifications are the norm.

1	(L v M) • (L v S)	
2	A → ~ L	
3	(~ S → L) → ~ M	∴ ~ A
4	L v M	Simp 1
5	L v S	Simp 1
6	S v L	Com 5
7	~ S → L	Imp 6
8	~ M	MP 3, 7
9	L	DS 4, 8
10	~ A	MT 2, 9

Note that much like in our truth tables, we again see two lines used to help organize our proofs. The vertical line is called a scope line. This visually shows us how far we are allowed to treat assumed statements as true. For now, the only assumed statements we have are the premises, so the vertical line is referred to as the **primary scope line**.



The horizontal line separates those assumed statements from the rest of the statements in our proof—the rest of these are *not assumed to be true*, so they must *all be justified*. If we are unable to provide a justification correctly and completely for all of the lines that appear below the horizontal line, our proof is not complete.

Many people enjoy doing proofs, because completing them is like doing little puzzles. They have a game-like quality to them, which is rewarding all by itself. However, the real reason we practice proofs is for the mental discipline they require. By showing us each step of a large inference, we are practicing the focus and patience needed to follow a complex train of thought. When we practice this, we are learning to use a linear problem-solving method that has far more application in everyday life than it might appear at first.

There's a Zen-like quality to proofs. We can learn the rules of logic for themselves, but over time you will realize that doing proofs offers many logic life lessons. Like any good meditation practice, the depth of insight is mostly up to the dedication of the practitioner. Logic is always useful in life…if you let it be.

Inference Rules of Logic

In order to read and understand the proof above, we need to be familiar with some basic rules of logic. The first set of rules we will learn are called "inference rules" because they justify small leaps we make from one set of statements to another. Later we will learn more rules to supplement these, but they work a little differently. The sample proof above makes use of both sets of rules, so it will be some time before you are able to fully read and understand it.

For now we will focus on the foundation of our system of natural deduction: the basic eight inference rules. The inference rules are not difficult to learn, but they require practice.

Each inference rule works on a type of statement. If you have ever played chess, then you know the drill. A type of statement is like a type of chess piece. Pawns have rules that apply directly to them. You cannot move a knight in chess the way you move a pawn. You cannot move a pawn the way you can move a queen. Each piece has rules that apply specifically to it.

The same is true of our inference rules. **They apply to certain *types of statements*.** So, a conditional statement has rules that only apply to it. Conditionals do not move like conjunctions (or any other type of statement). You cannot make "moves" in your mind with this type of statement in the same way you can move with a conjunction. If you try to move a bishop in chess the way you move a rook, you will be in violation of the rules of chess. The same applies with symbolic proofs.

In chess, we know the piece we are looking at by its shape (chess players, remember when you learned that "*the horsey looking thing*" is a knight). In symbolic derivations, we know the piece we have in hand by its main connective. The *main connective tells you* what type of statement is on any given line, and thus, *what type of rules apply to that statement*.

IMPORTANT NOTE: You must be able to identify the main connective correctly and quickly. If you cannot, you will not be able to do proofs (it's that simple). There is no way around this.

Metacognitive Checks

The rules of logic are useful for practicing the mental discipline of metacognition. Simply put, this is thinking about your own thinking. We will practice **checking our own thinking** by recalling simple elements of what we were thinking and **asking questions about it**.

> A mind that is skilled at metacognitive checks is *a superior thinker*: it is better at problem solving than a mind that relies on "insight" and "intelligence" to find solutions.

You don't have to be what some call "smart" to be good at doing logic; you do have to be *disciplined*.

The metacognitive checks we will perform help us understand if we are using the rules correctly. The first check is general. The second check will be specific to the rule itself and we will describe these as we look at each of them.

General Metacognitive Checks

At the most basic level, we check in with ourselves about our general use of the rule. We ask:

Q: Am I using this rule to accomplish the general task for which it was intended?

There are two general types of rules, divided according to the following general tasks:

- **Introduction Rules:** these "build up" and create a statement that you don't have but want to establish (think of stacking blocks to create a pyramid)
- **Elimination Rules:** these "break apart" a statement to gain access to a part of the statement or to another related statement (think of opening a cookie jar to get to its contents)

So each type of statement will have a specific way to build it up and a specific way to break it apart. Our first meta-check will be to verify that we are using the rule for its intended purpose.

Knowing when and where you will use these rules is knowing how to play the game. Every chess player worth their salt knows that learning the rules is only the beginning. Chess is not a game of rules; chess is a game of strategy. The rules are merely a way to define some structure in our possible moves; the game itself is *all about* the strategy.

Of course, the rules of chess need to be firmly understood. You do so first by memorizing the rules and then by putting them into practice. Simply put, you have to spend time playing the game. The same holds for symbolic derivatives: we learn the rules merely as a starting point. But do not be confused on this matter. Logic is not about rules. Logic is about using those rules to effectively guide your mind. *Logic (like life) requires strategy*.

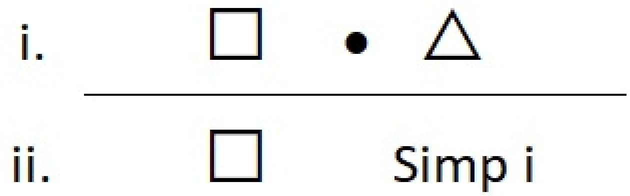
NOTE: In what follows, when we describe the inference rules, we will follow this convention:

- Provide the rule's *nickname* in quotes: this helps memorize the rule
- Provide the full *formal name* of the rule
- Provide the rule's *abbreviation* in parenthesis: for use in a proof
- Conjunction Rules

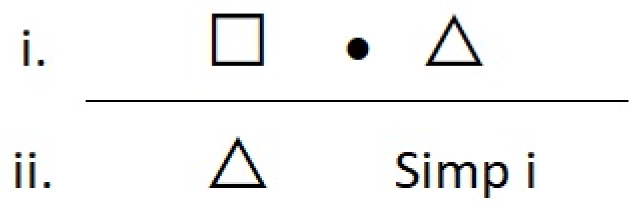
Conjunctions are fairly intuitive. If we know that an And statement is true, we know each one of its conjuncts is true. So that's our first rule.

"AND Elimination" Simplification (Simp.)

The rule of simplification is an elimination rule, so it breaks apart And statements. We use this when we have an And statement on a line and we want to take out one of its conjuncts to write it on a line by itself. The form looks like this:



We can also use Simp to break out the right-hand conjunct. Like this:



Meta-Check #1: The rule of Simp only applies to And statements, so the main connective of the statement on the line I reference must be a dot.

Meta-Check #2: The rule of Simp always requires *one and only one* line as the reference, so if I have more than one line referenced, I did something wrong.

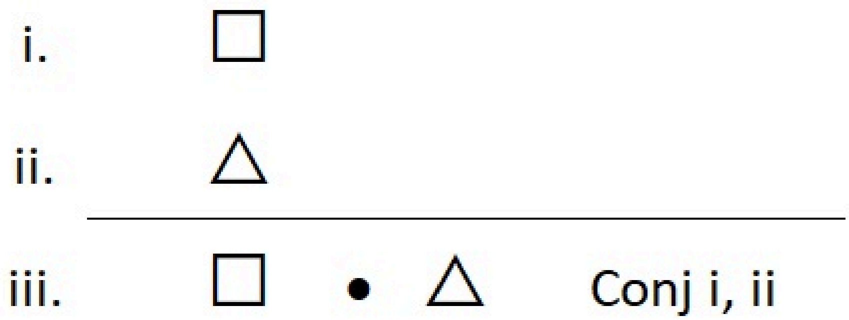
Note that in describing this rule, we did not number the lines as we usually do. This is simply a schematic of the form of the rule. In an actual proof we will often find that our references are spread out. For example, in our previous example we saw this line:

1	(L v M) • (L v S)	
2	A → ~ L	
3	(~ S → L) → ~ M	∴ ~ A
4	L v M	Simp 1
5	L v S	Simp 1
6	S v L	Com 5
7	~ S → L	Imp 6

We used Simp here to justify the truth of line 5 by reference to line 1. In what follows, we will use "i, ii, iii" etc. to describe the form of the remaining inference rules. Just keep in mind that the actual lines in a proof may be spread out very far.

"AND Introduction" Conjunction (Conj.)

The rule of conjunction is an introduction rule, so it builds And statements. We use this when we don't have an And statement, but we would like to assert one as true. To do so, we need to know that the truth of each conjunct has been previously established. The form looks like this:



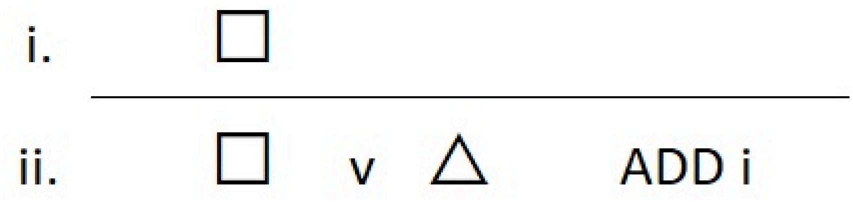
Meta-Check #1: The rule of Conj. only builds And statements. So, if the statement on the line I am justifying is some other type of statement, I did something wrong.

Meta-Check #2: The rule of Conj. requires *two lines* for justification. So, if I have only one line or more than two lines, I did something wrong.

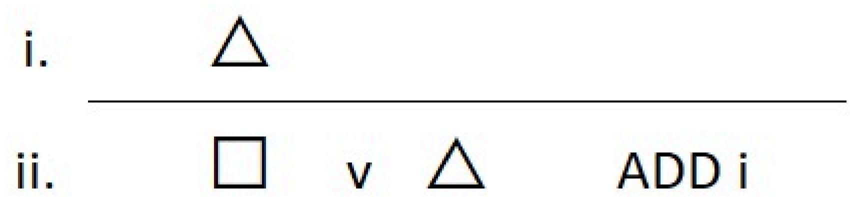
Disjunction Rules

"OR Introduction" Addition (ADD)

The rule of addition is an introduction rule, so it builds Or statements. We use this when we don't have an Or statement, but we would like to assert one as true. To do so, we need to know that at least one of its disjuncts is true. The form looks like this:



We can also use ADD to build Or statements if we know the right-hand disjunct is true. Like this:

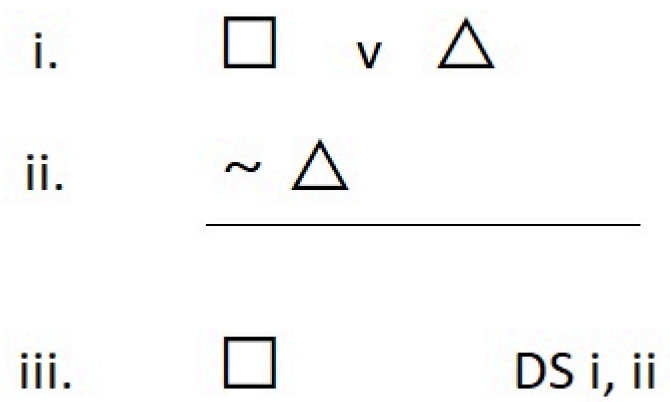


Meta-Check #1: The rule of ADD only builds Or statements. So, if the statement on the line I am justifying is some other type of statement, I did something wrong.

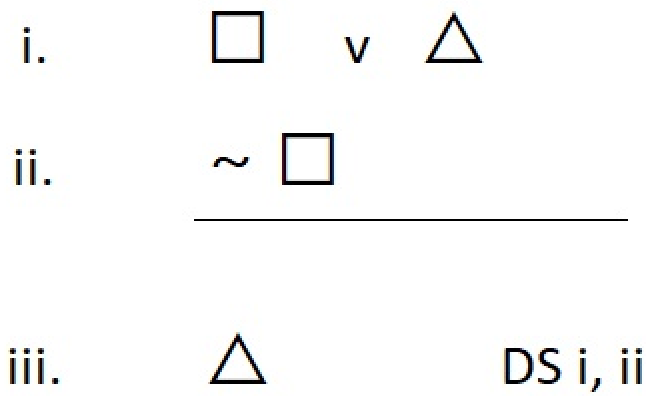
Meta-Check #2: The rule of ADD always requires *one and only one* line as the reference, so if I have more than one line referenced, I did something wrong.

"OR Elimination" Disjunctive Syllogism (DS)

The rule of disjunctive syllogism is an elimination rule, so it breaks apart Or statements. We use this when we have an Or statement on a line, and we want to take out one of its disjuncts to write it on a line by itself. The form looks like this:



We can also use DS to break apart Or statements if we know the left-hand disjunct is false. Like this:



Meta-Check #1: The rule of DS only applies to Or statements. So, if at least one of the lines I reference in my justification is not an Or statement, I did something wrong.

Meta-Check #2: The statement I am justifying must be one of the disjuncts that appears in the disjunction that I cite in my justification.

Meta-Check #2: The rule of DS requires *two lines* for justification. So, if I have only one line or more than two lines, I did something wrong.

Special Note on Double Negations

There are times when the use of DS (and other rules) requires that you provide the negation of a negation. For example:

I have ~ E v S on a line, and I want to break it apart to get the S

Since □ in ~ E v S is already a negation (i.e., ~ E), DS requires the negation of that **~** E. Under very strict systems, the inference rules require a ~ in front of whatever the □ happens to be, even in cases like this where it already contains a negation (such as this ~ E). That means we would need a **~ ~** E in order to use DS to get the S out of that line. These "double negations" are simply the result of a *hyper-strict* application of the inference rules.

Our system of natural deduction ***will not*** require "double negations" to justify the use of any rule which requires the negation of a negation. Our system will accept what we all likely already intuitively know. The negation of ~ E is in fact simply E.

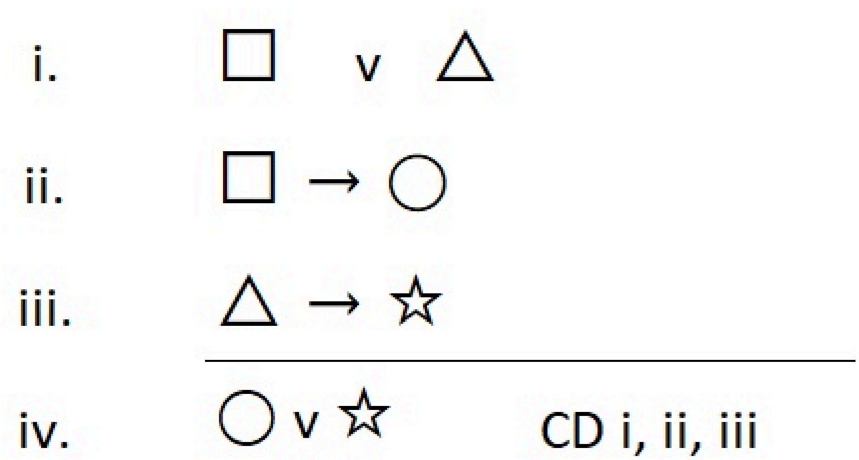
Of course, we will accept ~ ~ E as the negation of ~ E if we can provide that, but we may find it is easier to show that E is true. So we will also accept the following general rule:

The negation of ~ □ can be expressed as ~ ~ □ or it can be expressed as simply □.

"Big V OR Introduction" Constructive Dilemma (CD)

The rule of constructive dilemma is an introduction rule, so it also builds Or statements. We use this when we don't have an Or statement, but we would like to assert one as true.

Often, we use this after our attempts to use addition fail to build the Or we want. To do so, we need to know that three statements are true. The form looks like this:



Note that we used extra statement variables (◯ and ☆) to illustrate this rule.

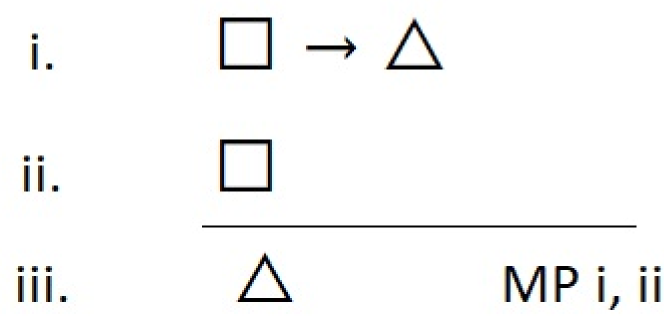
Meta-Check #1: The rule of CD only builds Or statements. So, if the statement on the line I am justifying is some other type of statement, I did something wrong.

Meta-Check #2: The rule of CD always requires *three* lines as a reference, so if I have fewer than that, I did something wrong.

Conditional Rules

"Conditional Elimination" Modus Ponens (MP)

The rule of modus ponens is an elimination rule, so it breaks apart conditional statements. We use this when we have a conditional statement on a line, and we want to take out its *consequent* to write it on a line by itself. The form looks like this:



Meta-Check #1: The rule of MP only breaks apart conditional statements. So, if at least one of the lines I reference in my justification is not a conditional statement, I did something wrong.

Meta-Check #2: The rule of MP always requires two lines to justify its use. So, if I have fewer than two lines referenced, I did something wrong.

Special Note: MP is always used to secure the tip of the arrow (the consequent). You can *never* get the antecedent out of a conditional statement. If you want that antecedent, you have to find another way to get it—the conditional will not give it up.

"Backwards Conditional Elimination" Modus Tollens (MT)

The rule of modus tollens is an elimination rule, so it also breaks apart conditional statements. We use this when we have a conditional statement on a line, and we want to write the *negation of its antecedent* on a line by itself. The form looks like this:

i.	□ → △	
ii.	~ △	
iii.	~ □	MT i, ii

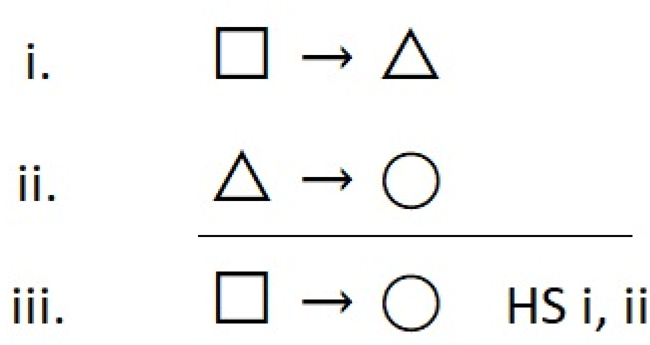
Meta-Check #1: The rule of MT only breaks apart conditional statements. So, if at least one of the lines I reference in my justification is not a conditional statement, I did something wrong.

Meta-Check #2: The rule of MT always requires two lines to justify its use. So, if I have fewer than two lines referenced, I did something wrong.

Special Note: MT is always used to secure *the denial* of the antecedent. Remember, you can *never* get the antecedent out of a conditional statement, but you can get its denial.

"Conditional Introduction" Hypothetical Syllogism (HS)

The rule of hypothetical syllogism is an introduction rule, so it builds conditional statements. We use this when we don't have a conditional statement, but we would like to assert one as true. Conditional statements build conditional statements; you will need two in order to make another. The form looks like this:

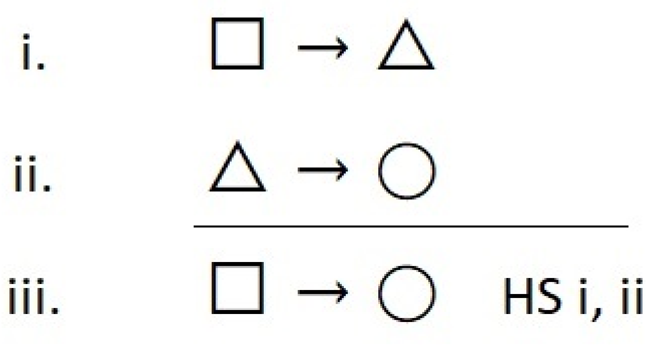


Note that we used an extra statement variable [i.e., ◯] to illustrate this rule.

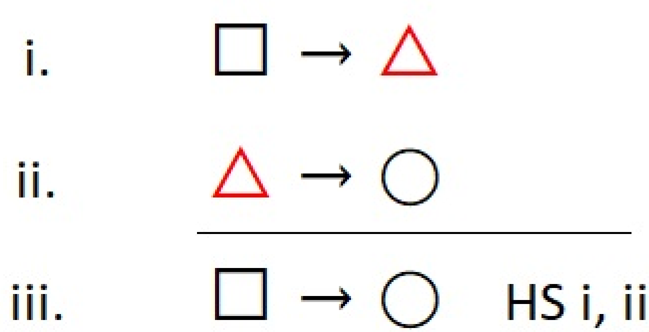
Meta-Check #1: The rule of HS only builds conditional statements. So, if the statement on the line I am justifying is some other type of statement, I did something wrong.

Meta-Check #2: The rule of HS always requires two lines to justify its use. So, if I have fewer than two lines referenced, I did something wrong.

Special Note: When using HS, the conditional statement you are trying to justify drives its use. The other two conditionals should have a very precise relationship to the one you are justifying. One should share the exact same antecedent, while the other should share the exact same consequent.



Additionally, each of these two conditionals should share the exact same statement that acts as a "middle term" (i.e., the △), which does not appear in the final conditional.



When we say "exact same," this means there should be no deviation at all.

Biconditional and Negation Rules

Note the following for biconditionals: We have no basic inference rules for making direct moves with biconditional statements. In the next chapter we will learn rules that give us robust access to these statements. For now, they are bricks—you can only use them in their entirety if they happen to be helpful in their current form for another inference rule.

Note the following for negations: We have seen that negations can be produced, i.e., they can be "introduced" through conditional statements (using MT). If you want to "build" or "introduce" a negation (i.e., you want a ~ ☐) that you do not have, then you need to start looking for a conditional statement that has that ☐ in its antecedent (i.e., ☐ → ∆).

However, if you wish to "break apart" a negation statement, you cannot currently do so with the basic 8 inference rules. We will have to wait until the next chapter to get rules with this power.

Using the Basic 8 Inference Rules

There are a few general restrictions on using the basic 8 inference rules. These are:

- Inference rules only apply to the main connective of a statement
  - e.g., You cannot apply an And rule to this statement: ~ R v (Y · P)
    - This is because that statement's main connective is the "v"
  - e.g., You can apply an And rule to this statement: ~ R · (Y v P)
    - This is because that statement's main connective is the "·"
- Inference rules must cite statements that appear above the line they are justifying
  - e.g., The justification for line 8 may only include lines 1-7
  - e.g., The justification for line 8 can never include lines 9 and higher
- Cited lines cannot include the same line they are justifying (i.e., they cannot be circular justifications)
  - e.g., The justification for line 8 can never include "8" in its citation

The savvy student will note that the first general restriction is the basis of some of our meta-cognitive checks. The other two general restrictions should be used as **additional general meta-cognitive checks**. You should always look to ensure that your references are neither circular nor provided as promises for what *you hope* will be true later in the proof.

Apart from these restrictions, the basic 8 inference rules are fairly easy to remember and use.

Strategy (a.k.a. Working Back from the Goal)

Allegedly, the second habit that highly effective people practice is to *begin with the end in mind*.[2] This is nothing more than starting your thought process with a clear sense of your goal, and then **working backwards** from there to make sensible progress towards it. Practicing symbolic derivations offers an excellent opportunity to discipline the mind in this habit—yet another logic life lesson.

To be fair, you *can* get through life (and get through many proofs) *without* being focused, *without* being strategic, and *without* being highly effective. You can flail about in a non-strategic manner, starting with what you have been given (in life or in the premises) and hope that something wonderful happens to get you to your goals. You *can* do this—whether or not you *should* do it or even want to do it is up to you.

Many people do not want to be highly effective, they just want to *get by*.

You are in charge of you. Maybe wonderful things will come your way…or maybe you will not wait for this: you will use your mind and develop a strategy to succeed.

The Two-Step Method

Approaching a proof in a strategic manner is actually very easy. There's a **simple two-step process** to help you develop a reasonable strategy to get you to your goal. It looks like this:

Step 1: *Identify* your goal

Step 2: *Identify* what you have to work with

Sounds simple, but two things make it hard when you are starting out.

First Challenge: You have to *identify* your goal and *identify* what you have to work with—you cannot get by simply "seeing" your goal or "reading" what your resources happen to be.

When we say "identify," we mean *understanding* the main connective of the goal and of the resources you have. For example, consider this statement:

B v ~ R

What *is* this statement?

If you answer that question by simply reading it, then you are not understanding it. If you say, "*I want B v ~ R*" then you are not understanding what you want. To understand it, you have to do more than point to it, and you have to do more than read it back to yourself. You have to look at the statement and recognize that *it is* an Or statement. You have to know the type of statement it is. Once you do that, you know what rules apply to it. Now you know *what possibilities there are* for what you want or what you have to work with in the proof. Now you have power over the statement.

2. See Dr. Stephen R. Covey, *The 7 Habits of Highly Effective People*.

Second Challenge: We lose focus very quickly.

You might know clearly what you want and what you have in hand. However, that clarity does not last. *Welcome to life*. Everyone loses focus more quickly than they are willing to admit. So, all that great power we just acquired over the proof will be gone in a split second.

The secret of this method is to *control your eyes*.

What you look at is only half the battle. The real secret is to control *when* to look at things. This is why we have a two-step method. Step 1 means put your eyes on the goal. DO NOT LOOK AT THE PREMISES. Step 2 means you are *now allowed* to put your eyes on your resources (the premises).

Virtually every student I have ever seen struggle has had a hard time doing proofs because they kept looking at the premises.[3]

They have a ***flawed*** *mental image in their mind: they believe that the premises have the answers*.

Put differently, they believe that the premises are valuable. They believe that the premises will unlock the path to the conclusion. This is false.

Or at least, we can say that focusing on the premises first is useful only when the proof (and life) is not very difficult. So, perhaps this is the most important logic life lesson: **your goals are what matter most, not your present resources**.

In what you have presently available (in the premises), you will find *confusion*.

In your goals, you will find *clarity* of purpose.

Thus, we'll start with the conclusion, with our purpose firmly in mind. The following cheat sheet (Inference Rules as Strategies [PDF, 130kb]) lays out the basic 8 inference rules for you to reference. Note that they are presented as strategic moves rooted in your goals, in what you want.

An Example of the Reverse Method

Let's walk through a sample problem to see how this method plays out. Consider the following problem:

1. (A v G) → K
2. K → (B → F)
3. A · B ∴ ~ M v F

We begin by extending our primary scope line far down the page and putting our conclusion at the bottom.

3. This has played out over two decades of classroom instruction, literally hundreds of times.

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	~ M v F	

Now we can take **Step 1**: Identify the goal.

Our Conclusion is an Or statement. This will either be contained in the premises in its entirety, or we will need to build it.

Now we can take **Step 2**: Identify what you have to work with

NOTE: By taking Step 1 seriously, we are in good shape to look at our resources—they will begin to make sense. Some things we have are valuable, some things are not. *All of this depends on a clear understanding of our goals*. This is the most important logic life lesson.

The key question we have right now is whether or not the conclusion is in the premises (in its entirety). So we look to the premises for this.

Q: Do you see it?

A: No

Q: What does that mean?

A: We have to build it. We need to “introduce” it.

Now we know we have to build it, so our questioning continues:

Q: How do you build it?

A: Depends on *what* you want to build...

Q: *What* do we want to build?

A: An **Or** statement (we know this *because we took the time* to identify the goal)

If you do not take Step 1 seriously, you have little to no knowledge of what you want to build. Almost every student who does not take Step 1 seriously struggles—and EVERY student who does not take Step 1 seriously struggles *more* than they would otherwise.

Knowing we have to build an Or statement, we can look over our rules to find suitable options. We are looking (a) for Or rules and (b) the rules that build them—the introduction rules. So we look and find two rules:

OR **Intro**duction: Addition (ADD)

"Big V" OR **Intro**duction: Constructive Dilemma (CD)

Between the two, we generally favor addition because it's fast and doesn't require much. That rule tells us we only need one thing to build an Or statement. Either disjunct will do.

So we return to **Step 1** to identify our new goal (or rather, to understand our two new options for the next goal):

Left-hand Disjunct: ~ M

Right-hand Disjunct: F

We don't need both—that's not what the rule of addition tells us. Addition tells us that our overall strategy will be to find one of these and use it to build our Or statement. One option is a negation and the other option is an atomic.

We can now take **Step 2** with a clear sense of what matters to us. We know what to look for in the premises.

A peculiar feature of addition *as a strategy* is that we should go look for both—we don't need both, but we should look for both, because if we find one right away, we may overlook an easier way to get the other one. *Look for both before you jump on the first one you see*. (Tip: This is another logic life lesson—be patient in developing your strategy.)

When we look at the premises, something odd jumps out at us. There are no ~ M statements anywhere. Heck, there are not even any atomic M statements! So that strongly suggests that our use of addition will require us to find the right-hand disjunct. At least there are atomic F statements in the premises. If we had an F on a line all by itself, we could build our conclusion off of it. So we write that down immediately above our main goal (i.e., the conclusion) as **a subgoal**. Like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	F	
	~ M v F	ADD _

So now **Step 1** begins again. We have something to look at and easily identify.

Q: What kind of statement is this?

A: An atomic

Q: Can we build atomics?

A: No. So, 99% of the time this means we must find them in the premises

Now we are empowered to look at the premises again. We know we are going to take **Step 2** again to look for compounds that contain what we want now.

Q: Where did we find that atomic F statement?

A: Premise 2

Q: Where is that F statement most immediately located?

When we ask about "*immediately located*," we mean to ask the following:

Q: Where does the atomic F appear as a *primary component* of a statement?

Sure, we find an F in the second premise, but the F is not a primary component of that premise. Premise 2 is a conditional statement. The main connective of premise 2 is the *first* arrow. When we look at the ☐ and ∆ we see that the atomic F is not the sole statement in either of those. In the ☐ we have "K" and in the ∆ we have a compound "B → F" statement. So the atomic F is not a primary component of the second premise; it is *buried inside* a primary component…and we need to get it out.

The atomic F is located *in the consequent* of the second premise. *Inside* there, *it is* a primary component of that little conditional. So we can say that the atomic F is most immediately located in that little conditional statement. (Put differently, if we look only at that statement, we see that the F is the consequent of this little conditional statement.)

(B → **F**)

Life would be great if we could get this statement on a line all by itself. After all, if this were on a line all by itself, we could apply our rules to it to get what we want.

Remember: the inference rules only apply to the main connective of a statement

Having this little conditional on a line by itself means that one of our conditional elimination rules can be used to break it apart. So we write that statement down on a line immediately above our last subgoal. Like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	B → F	
	F	
	~ M v F	ADD _

We now can see pretty clearly that our F will be secured by breaking apart the B → F that we are saying we hope to get.

As we noted earlier, breaking B → F apart will require a conditional elimination rule, one that delivers to us the statement in the Δ position of the conditional. We look over our cheat sheet to find the rule that will break apart a conditional and give us the Δ. This rule is called modus ponens (MP). Now we have a strategy for getting the F: we know we'll try to use MP to justify it. So we write the rule down to remind ourselves of the plan:

1	$(A \vee G) \rightarrow K$	
2	$K \rightarrow (B \rightarrow F)$	
3	$A \bullet B$	$\therefore \sim M \vee F$
	$B \rightarrow F$	
	F	MP _, _
	$\sim M \vee F$	ADD _

Note that we also drew out how many lines a correct use of MP will take. This helps with our meta-cognitive checks. If we cannot fill in both of these lines, we will know we did something wrong.

Great! We have a plan to get the F, but it will require the B → F statement on a line. This is our newest subgoal. So…Step 1: identify it (it's a conditional); Step 2: take stock of your premises.

We see that we had this statement (in its entirety) tied up in a larger statement on line 2. Line 2 is also a conditional statement—and it has what we want. So we need to break line 2 apart. We review our conditional rules and once find again that modus ponens will give us the consequent. So that's our strategy, and we note it like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	B → F	MP 2, _
	F	MP _, _
	~ M v F	ADD _

Notice here that when we wrote down MP, we included the direct reference to line 2. We know we'll need to use MP to break apart that specific line, so we can write it down to finish half of that justification. MP takes two lines as a justification, so we still have a blank spot in the justification. We don't know what that line will be yet, so we just leave it as a underscored line for now.

Pausing for a moment here, we see that a plan is emerging for how to get our atomic F statement. The F now appears as the primary consequent of this line. So modus ponens will allow us to break apart this statement to get the "F" statement. Modus ponens is being used twice to deliver to us two different consequents.

Since MP requires two lines, this rule gives us clear insight into what we need to proceed. Our application of MP to get the atomic F will reference the line in which B → F appears, as well as another line. The rule of MP tells us that we need this conditional, and we also need the ☐, the antecedent. In this case, that ☐ contains the atomic "B." So we write this down as a subgoal. Like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	B	
	B → F	MP 2, _
	F	MP _, _
	~ M v F	ADD _

However, in order to get that "B → F" on a line by itself, we have to wrestle it out of the second premise.

2	K → (B → F)

MP tells us we *need* to secure the □ of the second premise in order to get the "B → F" alone. In this case, we require an atomic K to pay off the conditional in the second premise. So we write that subgoal down too. Like this:

1	$(A \vee G) \rightarrow K$	
2	$K \rightarrow (B \rightarrow F)$	
3	$A \bullet B$	$\therefore$ $\sim M \vee F$
	K	
	B	
	$B \rightarrow F$	MP 2, _
	F	MP _, _
	$\sim M \vee F$	ADD _

Now we can return to **Step 1** for these subgoals. Identify the first.

B is an atomic statement

This likely appears in the premises somewhere

Step 2 repeats: Identify what you have to work with. We look at the premises now:

1	$(A \lor G) \rightarrow K$
2	$K \rightarrow (B \rightarrow F)$
3	$A \bullet B$

We find an atomic B in the third premise. Even better, we find the atomic B as one of the primary components of line 3. So we identify what kind of statement line 3 is so that we know how to break it apart.

Line 3 is a conjunction

Simplification (Simp) is the conjunction elimination rule

Because Simp requires nothing more than a conjunction as the main connective, we can directly apply it to the third premise. Our justification for the atomic B is complete. Like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	K	
	B	SIMP 3
	B → F	MP 2, _
	F	MP _, _
	~ M v F	ADD _

Great. Now we return to **Step 1** for our remaining subgoal.

K is an atomic statement

This likely appears in the premises somewhere

Now on to Step 2 to look for this in the premises. We find an atomic K tied up in a compound statement on line 1:

(A v G) → K

When we identify this statement, we see that:

Line 1 is a conditional statement

Modus ponens is the conditional elimination rule that breaks apart conditionals and yields the consequent

So we have a strategy in place to secure our atomic K statement. Remember, we write that justification down and indicate how many lines it requires. Like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	K	MP 1, _
	B	SIMP 3
	B → F	MP 2, _
	F	MP _, _
	~ M v F	ADD _

We know one of these references is the first premise. We don't need to prove that line 1 is true; we directly invoke it in our justification. So we filled that direct reference in and included an underscored line to remind us that we need another statement to complete the MP.

Remember that MP tells us that we have to provide the conditional on line 1 with its antecedent.

(**A v G**) → K

So we know we need this new subgoal and we should write it down like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	A v G	
	K	MP 1, _
	B	SIMP 3
	B → F	MP 2, _
	F	MP _, _
	~ M v F	ADD _

Now **Step 1** begins anew. We identify this new subgoal "A v G" as a disjunction.

Now Step 2 repeats. We identify if we have this statement in its entirety in the premises.

We do not

So that means we have to build it

Addition is the quickest fastest way to build Or statements

We know what addition requires: *just one* of the disjuncts

So we *look for both* (the atomic A and the atomic G)—in case one is easier to secure than the other

In this case, it seems there are no other atomic G statements in the premises. So this strongly suggests that we will build the A v G by securing the other disjunct, the atomic A. So we write this down as our next subgoal (along with the ADD justification for our "A v G"), like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	A	
	A v G	ADD _
	K	MP 1, _
	B	SIMP 3
	B → F	MP 2, _
	F	MP _, _
	~ M v F	ADD _

Step 1 takes place for "A." We identify this as a simple atomic statement, and that means it is likely in the premises somewhere.

Step 2 takes place. We look for the atomic A and find it in the third premise. As we saw earlier, this is a conjunction, and the A appears as the primary conjunct in it. So this means we can use simplification to break the third premise apart and get what we want. So we write our justification like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
	A	SIMP 3
	A v G	ADD _
	K	MP 1, _
	B	SIMP 3
	B → F	MP 2, _
	F	MP _, _
	~ M v F	ADD _

Notice now that we seem to have everything we need. We have a plan for everything and that plan does not require any further statements. At least, this is how it seems. **We will perform a *meta-cognitive check* when we attempt to fill in the rest of our justifications.** To do so, we start by filling in the numbering for each of our lines. Like this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
4	A	SIMP 3
5	A v G	ADD _
6	K	MP 1, _
7	B	SIMP 3
8	B → F	MP 2, _
9	F	MP _, _
10	~ M v F	ADD _

Now we have numbers to use in the rest of the blank spots in our justifications. If we find we are unable to properly cite a line by filling in all the blank spots, our **meta-cognitive check** will have revealed that we are not yet done. So we give it a try and find this:

1	(A v G) → K	
2	K → (B → F)	
3	A • B	∴ ~ M v F
4	A	SIMP 3
5	A v G	ADD 4
6	K	MP 1, 5
7	B	SIMP 3
8	B → F	MP 2, 7
9	F	MP 7, 8
10	~ M v F	ADD 9

In this case, we were able to provide a complete and correct citation for each rule we used. This completes the justifications for each line in our proof.

A Word on Doing Well

This reverse method is not always intuitive. You will need to practice it to grow comfortable with it. You are well advised to do so even on easy problems.

Easy problems can often be done "quicker" by proceeding from the premises down to the conclusion, using a forward method. *This is a trap!* You will get the mistaken impression that this is better, simply because you are meeting with early success on *easy* problems. You are being fooled by your own success.

As problems become increasingly challenging, you will need the reverse method more and more to quickly dispatch them. Your "forward method" will become increasingly slower and harder to pull off. So don't be fooled.

By practicing the reverse method even on easy problems, you get comfortable with the method quickly. When the harder problems appear, your comfort and familiarity with strategic thinking will pay off.

Additionally, in the next chapter we will look at replacement rules. When these arrive, the floodgates will open and problems will become far more challenging. You will need a firm foundation you can count on. The forward method becomes a recipe for wasted time and confusion. The reverse method will be a reliable base that you can build replacement rules on without becoming overwhelmed.

Here is a useful website to try if you are working with pen and paper. This is not a replacement for meta-cognitive checks, but it is a nice resource to enter problems into as a final verification.

Proof Checker (https://proof-checker.org/)

NOTES:

- You must enter a proof into this system (all Ps and the conclusion).
- Additionally, you must use *their* annotation system (carets ^ for dot ·, etc.).
- They require double negation
- Not all rules are shown in the sidebar of rules (e.g., DeM is not listed, but the system accepts it as a proper justification)

Practice Problems

Please see section **8.1** of the Power of Logic Web Tutor for a selection of problems using only the first eight inference rules.

These can be done directly online or printed out for pen and paper practice.

Note that sections 8.2 through 8.6 of the Power of Logic Web Tutor provide practice problems that make use of advanced rules. We will cover these rules in the next chapter of this textbook, and you may wish to revisit these sections of the Web Tutor for additional practice.

8.

ADVANCED PROPOSITIONAL LOGIC

Introduction

Our basic 8 inference rules allow us to demonstrate the validity of many arguments. However, they do not allow us to do so for all formally valid arguments. There is only so much we can do within our system thus far, so we will need additional rules that open up strategic possibilities.

Replacement rules are also referred to as "equivalence rules," because they express logically equivalent statement forms. By using equivalence rules, we are simply exchanging one statement form for another, logically equivalent, statement form. We are not inferring something new. We are simply "restating" what was already established.

You might wonder why we would do this. The answer is found in everyday life.

> If you have ever stood in front of a vending machine with a $20 bill in your hand, you know the frustration of having what you need but having it in a difficult form. What to do? Of course, you go ask someone to trade you for a more useful version of your money. You will be happy to get a $10 bill, a $5 bill, and five singles. You haven't made money, you haven't lost money. You got the exact same value.

The same is true of our equivalence rules. We don't lose or gain value—the truth value of our statements must stay the same to make it a fair trade.

Now that I have my $20 of value in a new form, I have greater ability to do things *I want to do* with it. Here again we see the emphasis on the goal. We use the replacement rules wisely when we know what we want. Knowing our goals makes their use sensible—it would be very odd indeed to walk around randomly asking people to break your $20 bill.

So, nothing has fundamentally changed with our method. We still start with our goals. We still use the reverse method of generating a linear problem-solving strategy to achieve those goals. The only difference is that now we have many more strategic opportunities. After all, when we identify our goal and we identify our resources, we are identifying a type of statement. Now, those types (i.e., those $20 bills) can morph into different types of statements. This greatly expands the range of moves we can make. We gain tremendous flexibility in how we solve a proof as well as the ability to solve proofs that were previously closed off from us. This is very good. However, this also adds a great deal of complexity to our proofs. *Now the really fun proofs can begin!*

We will look at 10 replacement rules overall, but we'll do so in two sets. The first set is a good introduction to using replacement rules. The patterns there are not too difficult to learn. The second set is a bit more complicated.

We will also introduce a new symbol:

::

This symbol expresses equivalency (think of it like a mathematical = sign). The four dots are different from the three dots that expressed an inference. *Three dots are directional*. You can only infer the conclusion on the basis of the previous statement(s). Consider:

> The rule Simp tells us that knowing ☐ ·∆ allows us to infer that ☐ is true. However, knowing that ☐ is true *by itself* tells you nothing about the truth of the whole conjunction ☐ ·∆. You cannot infer that statement on the basis of the conclusion.

With the ∷ things are different. You can assert that one side of the ∷ is true on the basis of what is on the other side. *Four dots are bidirectional*.

First 5 Replacement Rules

Double Negation (DN)

Double negation is a straightforward rule, and as such is a good place to start. The form of the rule is as follows:

☐ ∷ ~ ~ ☐

We see that this rule applies to any statement of any form. Any ☐ can be replaced with its equivalent version ~ ~ ☐. Your English teacher may not approve, but in our symbolic language this is a perfectly acceptable statement form. In some cases, this is exactly the form that is required to satisfy the use of a different rule (think: an instance of modus ponens in which the antecedent of the conditional is something like ~ ~ G, and you only have a G to work with).

DN is such a simple rule that it is also useful in illustrating some unique features and requirements of the replacement rules. These are:

1. You can use them anywhere in a statement that the patterns apply (i.e., their use is not restricted to the main connective)
2. You can only use them one at a time
3. Replacement rules *always only* reference one line
4. You can use them on the same statement in a different part of the statement (so long as the pattern they express applies)

So let's look at a simple example. Consider the following:

8	P	MP 1, 5
9	~ ~ P	DN 8
10	~ ~ P → J	SIMP 3
11	J	MP 9, 10

I used DN to change line 8 into the ~ ~ P that I needed to satisfy the conditional on line 10. However, I didn't need to use DN in this specific way. Eventually I want to get the J out of that conditional, but I could instead have done this:

8	P	MP 1, 5
9	~ ~ P → J	SIMP 3
10	P → J	DN 9
11	J	MP 8, 10

Notice that in this use of DN, I only applied it to the antecedent of the conditional. Because the ∷ of all replacement rules is bidirectional, I can use DN to take away a ~ ~ if I want to do so. This "cleans up" the conditional and changes what is needed to deliver the J that I want. Either use of DN is permissible, and the decision in cases like this comes down to personal preference.

Keep in mind that I am not restricted to using DN on atomic statements. We can use it on any □ we elect, even if it is a compound statement. For example:

8	~ ~ (C v P) → H	SIMP 5
9	C v P	SIMP 3
10	~ ~ (C v P)	DN 9
11	H	MP 8, 10

We can use DN on an entire line or on just part of a line. Put differently, I can use DN on the main connective of a line or on just a component of the line.

The one thing we cannot do with any replacement rules is use a rule more than once on a given line or use it with other rules *on the same line*. If I want to make multiple changes, I need to do so in two different steps. Like this:

8	~ ~ (C v ~ ~ P) → H	SIMP 5
9	C v P	SIMP 3
10	~ ~ (C v P)	DN 9
11	~ ~ (C v ~ ~ P)	DN 10
12	H	MP 8, 11

Note the red emphasis placed on each application of DN to illustrate each single application of our replacement rule. Also note that in my second application of DN (on line 11) I applied the rule *to line 10*. I did not apply DN "again" to line 9. This is important because any time I apply a replacement rule to a component of a given line, the other parts of that line do not change. When we look at line 11 we see that *all other aspects of line 10 are still there*—only the single use of DN has changed line 10 into line 11.

A Word of Warning:

You can see how flexible the replacement rules can be…and how much trouble they can get you in if not used with purpose. In theory, you can make an infinite number of moves with just this one replacement rule. If you are not careful, you will quickly go down a deep rabbit hole of desperate replacements…hoping something *wonderful* happens.[1]

Commutation (Com)

The rule of commutation applies to And statements and Or statements. When we look closely at the basic conditions of truth for these statements in a truth table, we see that the order of the component statements

1. Perhaps here is a good place to note that the play on words is not by accident. Lewis Carroll famously eschewed logic in Alice in Wonderland. Without logic to focus our efforts, we really do get lost down a hole of endless possibilities. This may be great for finding your way through Wonderland (a world of pure fantasy), but it is a particularly bad way to find your way to your goals. Yet another of logic's real-world life lessons.

has no impact on their truth. So this rule allows us to make use of that knowledge. The form of the rule is as follows:

□ · △ :: △ · □

□ v △ :: △ v □

Commutation allows us to swap the order of the conjuncts or disjuncts on any subsequent line of a proof. Here's an example:

8	(C v P) → H	SIMP 5
9	P v C	SIMP 3
10	C v P	COM 9
11	H	MP 8, 10

You might think this is a trivial rule, but it has great use, especially when combined with the following rule.

Association (As)

The rule of association is used like a rule to signal friendship. Close friends are bound together by ( ) in a large compound. Consider this statement:

S · (N · K)

Who are the besties?

And who is left out in the cold?

Clearly we see that (N · K) are best friends and S is kind of the third wheel. But sometimes we can change that. The form of association is as follows:

□ · (△ · ○) :: (□ · △) · ○

□ v (△ v ○) :: (□ v △) v ○

Notice that the order of the components does not change. All we are doing is changing the grouping of the statements. If we want to change the order of the component statements, then we have the rule of commutation for that. A combination of association and commutation gives us great control over how such statements can look. We will use these two rules together often.

A Word of Warning: We can only do this when we have a string of ·s or a string of **v**s. This rule *does not work* with any other type of statement.

DeMorgan's Law (DeM)

This is one of the most useful replacement rules we have in our system. DeMorgan's Law will become a great personal friend, as it will get you out of trouble many times. The rule itself expresses what the denial of an And (or an Or) statement really means. The form looks like this:

~ (□ · Δ)   ::   ~ □ v ~ Δ

~ (□ v Δ)   ::   ~ □ · ~ Δ

The savvy student will notice two things. First, we are not simply moving the ~ inside the ( ) when we use this rule. Doing this does not produce logically equivalent statements. When we move the ~ inside the ( ), *we must also change the operator* there. The "·" becomes a "v" and the "v" becomes a "·".

The savvy student also notices something else. *Having* the denial of an Or statement now becomes very fruitful (e.g., when you have one given to you in the premises). We can transform that into an And statement, and Simp makes it easy to grab whichever side we want.

Additionally, we might be frustrated if we *need* the negation of an And statement (e.g., when you need that to pull off a modus tollens or a disjunctive syllogism). At least, we are only frustrated until we realize that all we really need is the negation of either side. With a ~ □ or a ~ Δ on a line, we can use Add to build the rest and flip back with DeMorgan's into the denial we wanted all along.

Identifying these patterns is key:

"*Negation of an And statement*"

"*Negation of an Or statement*"

When those phrases pass through your mind, you will have a good chance of remembering this rule and the power it has to get you to your next move.

Also note that as with all replacement rules, we can use this anywhere within a statement, on the main connective or on just a component of the statement. For example:

13	K → ~ (B v F)	SIMP 2
14	K → (~ B • ~ F)	DeM 13

Contraposition (Cont)

The rule of contraposition applies to conditional statements. The form looks like this:

□ → Δ   ::   ~ Δ → ~ □

The savvy student may recognize the relationship that this rule expresses. Ask yourself: What does this remind you of when you read it from left to right?

Go ahead, take another look…

We'll wait…

If you said, "*Hey, it looks kind of like modus tollens,*" you are right. Contraposition expresses the same relationship found in all conditionals; it just doesn't allow us to infer that ~ ☐ is actually true. Instead, the rule tells us what to expect if the necessary condition is false (i.e., the sufficient condition could not have been met).

All this rule allows us to do is swap the positions (thus, *contra*–position) of the antecedent and consequent. Of course, we cannot do this cleanly. If we just swapped positions the way we do with commutation, we would get a false equivalency. Swapping positions in a conditional requires us to *deny each component* in the new conditional.

This rule is very useful for setting up a hypothetical syllogism such as this:

5	~ L → S	SIMP 3
6	(D v R) → ~ S	SIMP 1
7	~ S → L	CONT 5
8	(D v R) → L	HS 6, 7

This rule is also a favorite of students who (for their own personal reasons) just don't like to use modus tollens. Now you don't have to put up with MT if you don't want to use that rule. You can always convert a strategy that would require the use of MT into a strategy that uses the more comfortable MP rule.

8	(C v P) → H	DS 4, 6
9	~ H	SIMP 3
10	~ (C v P)	MT 8, 9

8	(C v P) → H	DS 4, 6
9	~ H	SIMP 3
10	~ H → ~ (C v P)	CONT 8
11	~ (C v P)	MP 9, 10

Of course, what you need to pay off the new conditional is *exactly the same* as what you would need to pull off the MT (note the red emphasis placed on ~ H to point this out). The proof does not become easier in this regard, but at least you'll feel better about it. That's something.

First 5 Replacement Rules Problem Set

First 5 Replacement Rules Problem Set [PDF, 41 KB]

Second 5 Replacement Rules

Redundancy (Re)

We claimed that the second set of replacement rules would involve more complex patterns. This is mostly true—except for this one. This is likely the most banal rule we have, but it serves a useful purpose from time to time. Here is the form:

□ :: □ v □

□ :: □ · □

Note that this rule applies to any statement, i.e., any □. The use of the rule comes into its own when we find odd situations where we may be able to produce an Or statement like □ v □, but cannot produce the □ on its own (think: being forced into a constructive dilemma because you couldn't pay off either conditional to use MP).

6	B v ~ L	DS 1, 2
7	B → Y	SIMP 3
8	~ L → Y	MP 2, 7
9	Y v Y	CD 6, 7, 8
10	Y	RE 9

Redundancy allows us to fold this into the □ we wanted all along.

Distribution (Dist)

The rule of distribution has two applications, one for *peculiar* conjunctions and one for *peculiar* disjunctions. The form looks like this:

□ · (Δ v ○) :: (□ · Δ) v (□ · ○)

□ v (Δ · ○) :: (□ v Δ) · (□ v ○)

Note that what we see on the left side of the four dots are the And and Or statements with unique features. The And statement is peculiar because it contains a conjunct which is itself an Or statement.

Alternatively, in the second application of the rule, the peculiar Or statement contains a disjunct which is itself an And statement. When we see these patterns, we can distribute the other conjunct (or disjunct) *along* *with the main connective* into the conjunct which is an Or (or, in the second application, the one which is an And statement). What was once a minor connective becomes the main connective. See here:

☐ · (Δ v ○) :: (☐ · Δ) v (☐ · ○)

☐ v (Δ · ○) :: (☐ v Δ) · (☐ v ○)

Another way to look at this rule is to start from the right side of the four dots. Focus on that side and ask yourself this:

Q: What is "odd" or "funny looking" about the statement form on the right?

If you noticed a repeated presence of the ☐ claim, then you have a good chance of recognizing this pattern.

(☐ · Δ) v (☐ · ○)

(☐ v Δ) · (☐ v ○)

We see here that the exact same statement (☐) appears in both conjuncts (or disjuncts).

That's odd. How did it get there?

A: It was distributed in.

So that means (since it's four dots) that we can un-distribute it and isolate it. That's what we're seeing to the left of the four dots.

This is very handy when we want that ☐. After all, if the result is an And statement (see the first application), then we're going to Simp(ly) grab it after the distribution.

NOTE: In this rule, the *order of the component statements* (the ☐ Δ ○) does not change. Again, if we wish to switch the order of these statements, we have commutation to do that.

Exportation (Exp)

The rule of exportation applies to peculiar conditional statements. The form looks like this:

(☐ · Δ) →○ :: ☐ → (Δ → ○)

Exportation might look a bit confusing at first, but in our everyday lives we frequently find ourselves reasoning in this way. Look at the right-hand side of the four dots. You might recognize this pattern from a road map. If this were a trip we were planning, we might ask:

Q: How do you get to the town of ○?

A: You have to go through the town of Δ.

Q: Is that all?

A: No, you also have to go through the town of ☐.

Moral: You have to go through the towns of ☐ *AND* Δ.

So when we look at the left side of the four dots, this is exactly what we're saying. You have to go through the towns of ☐ · Δ to get to the town of ○.

An alternative way to think about exportation is in the name of the rule. Looking at the left side of the four dots we see that the antecedent is a compound—it's an And statement. There are two things in the antecedent, so we can "export out" the right-hand conjunct (the △) into the overall consequent. The consequent then becomes its own little conditional with that same △ as its consequent (△ → ○).

A Word of Warning: Use care in how you think about what you see in this rule. On the *left side* of the four dots is a conditional whose ANTECEDENT is an And statement. We don't care what the consequent is on this side of the four dots, and we cannot use this rule if the ANTECEDENT is a Conditional statement. We can only use it if there's an And statement in the antecedent.

Also, on the *right side* of the four dots is a conditional whose CONSEQUENT is itself a conditional. That's all that matters here. We don't simply see any ol' string of conditionals—we see a very specific pattern. After all, the following might be rightly called a "string of conditionals" in a sense:

(☐ → △) → ○

However, you cannot use exportation on this form. Having a conditional in the antecedent doesn't help. The little conditional we see in "a string of conditionals" *has to be* in the consequent.

Material Implication (MI or IMP)

The rule of implication applies to conditional statements (and to disjunctions). The form looks like this:

☐ → △ :: ~ ☐ v △

Implication expresses the basic concept of bivalence as it applies to the antecedent of conditional statements. That antecedent is either true or it is false. If it is true, then we know that △ follows. If it is false, then we really don't know anything except that it is false. So either ~ ☐ is true or △ is true. Simple.

Implication is an extremely powerful rule. IMP enables us to transform *any* conditional into an Or statement (and vice versa). So that means every strategy we have for proving Or statements is now applicable to proving conditionals (and vice versa).

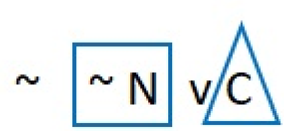
We should also note that the right-hand side of the four dots is a bit misleading. To the right, we first see a peculiar looking Or statement. This Or's left-hand disjunct is a negation, and that's a tip-off that IMP may apply. However, this does not need to be the case. After all, consider this:

N v C

Can we apply IMP to this? Of course, we just need to "see" the negation hiding here. The rule of Double Negation makes this easy:

~ ~ N v C

Now we have a statement which fits the form of the right-hand side of the rule.



The result of using IMP on this would thus be as follows:

~ N → C

The moral of that story is that in truth, ANY conditional can be exchanged for an equivalent Or statement, and ANY Or statement can be exchanged for an equivalent conditional statement. You don't need to actually have the negation present, as we saw through the use of DN to produce it. The shorthand rule for using IMP is the following phrase:

Negate the antecedent, leave the consequent alone.

Alternatively, you could say:

Negate the left-hand disjunct, leave the right alone.

Always leave the right-hand side of the statement you are Implicating alone. This will make applying IMP to any conditional or Or easy, regardless of whether an actual "~" appears or not.

Material Equivalence (ME or Equiv)

The rule of Implication applies to biconditional statements. There are two applications of the rule. The forms look like this:

(☐ ↔ △) :: (☐ → △) · (△ → ☐)

(☐ ↔ △) :: (☐ · △) v (~ ☐ · ~ △)

Finally! We have ways to access biconditional statements! There are two versions of this rule, so let's nickname them to help memorize the form:

The "Name" version: Anything which is a "bi" has "two" in it. In this case, a "*bi*-conditional" has two conditionals buried in it. So we're just unpacking those two conditionals. Since we really do have *both* of these, the two conditionals are joined with a *dot*.

(☐ → △) · (△ → ☐)

The Truth Table version: If you remember the basic conditions of truth for biconditionals, you know that there are two ways a biconditional can be true. Since we are assuming this biconditional is true, we can say that *at least one* of these conditions is true. This version of the rule simply expresses those conditions, and we use a "v" to join them because we don't know which of these two conditions holds. It could be one *or* it could be the other.

(☐ · △) **v** (~ ☐ · ~ △)

With material equivalence, we now have a way to use every biconditional as a conditional. We'll just unpack it with ME and then Simp out the one we need. The truth table version is handy when it comes time to prove a biconditional is true. After all, we really just need to prove one of the disjuncts, Add the other, and then use ME to assert our biconditional.

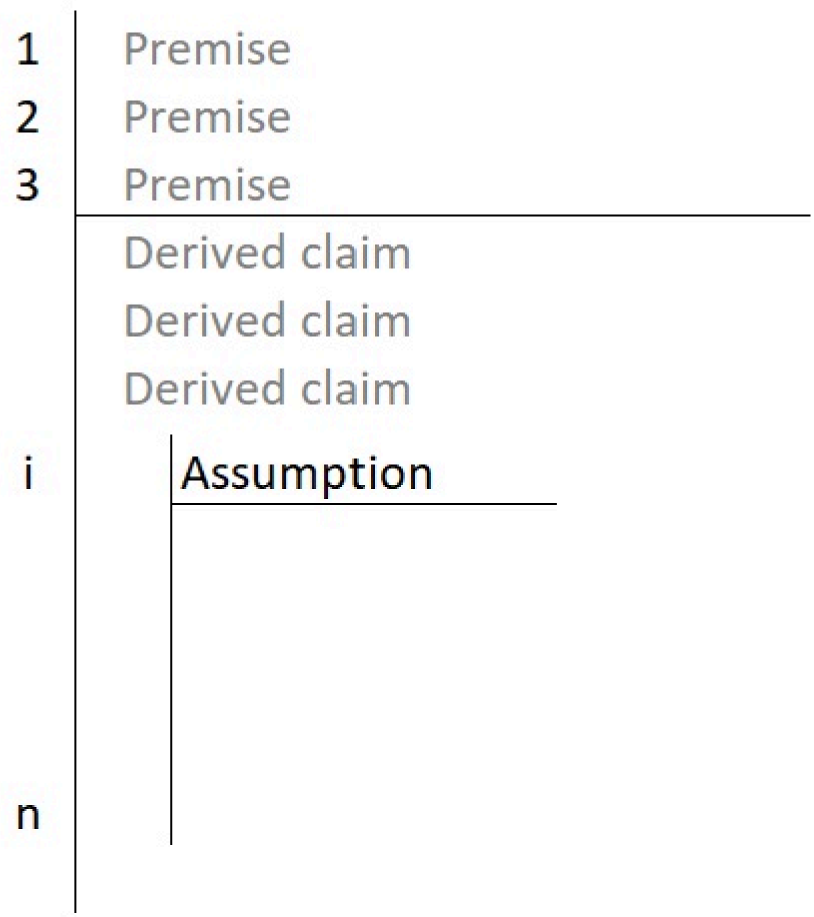
Second 5 Replacement Rules Problem Set

Second 5 Replacement Rules Problem Set [PDF, 40 KB]

Sub-Proof Methods

The next two "rules" are more akin to modifications of the existing methods we have already learned. Now, rather than passively accept the range of the primary scope line, we will actively *initiate additional scope lines* with their own limited ranges. Like the primary scope line, these secondary scope lines define a range under which statements can be treated as true.

The initial claim that starts the process is not being asserted as true—it is merely assumed for the duration of the scope line. The savvy student recalls that this is exactly the same as how our premises are treated—they are not proven to be true, they are merely assumed to be true for the purpose of the proof. So, schematically it looks roughly like this:



We have two reasons to initiate an assumption with its own scope line. One is to help us prove conditional statements. The other is to help us prove that an "assumed statement" cannot possibly be true. These two reasons are formally expressed in the following two sub-proof rules.

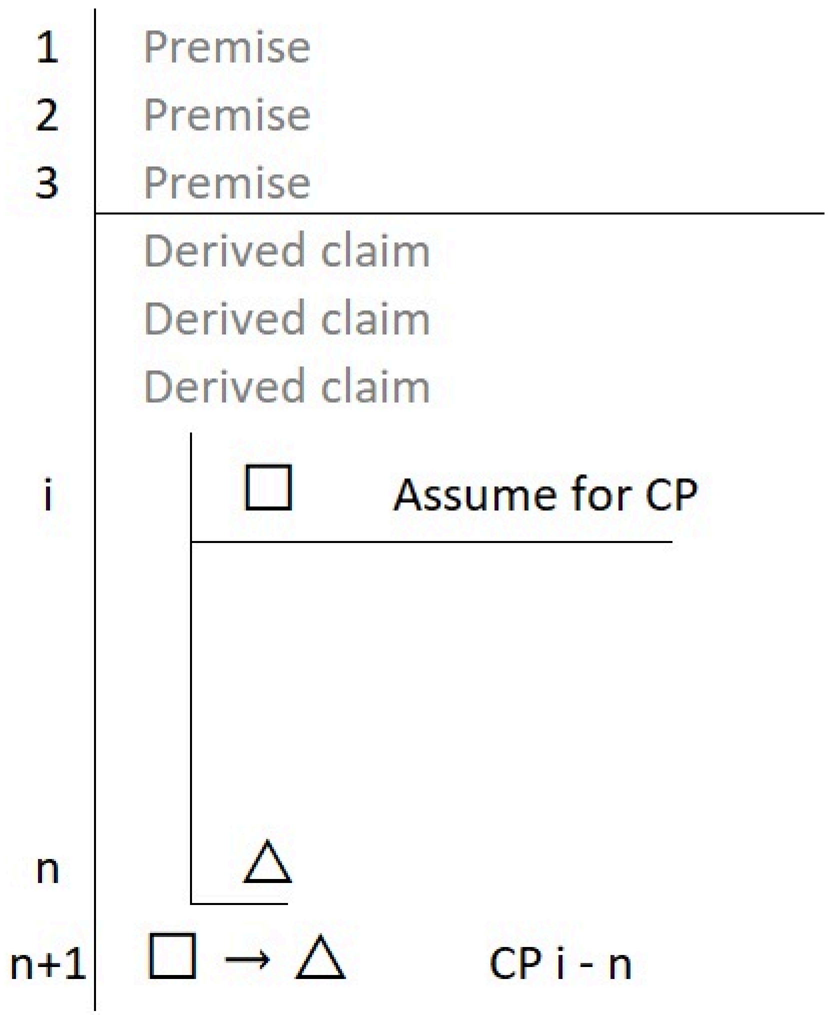
Conditional Proof (CP)

This rule builds conditional statements; it is a rule tailor-made to do one and only one thing: *build conditionals*. To use it, we initiate an assumption and draw out our secondary scope line. We then proceed as usual in a proof. The objective is to make our way down to the consequent of the conditional we want to prove is true.

Thus, the basic setup of a conditional proof is driven by the conditional we want to prove is true.

- Assume that the *antecedent* is true
- Then demonstrate that the *consequent* is true under this assumption

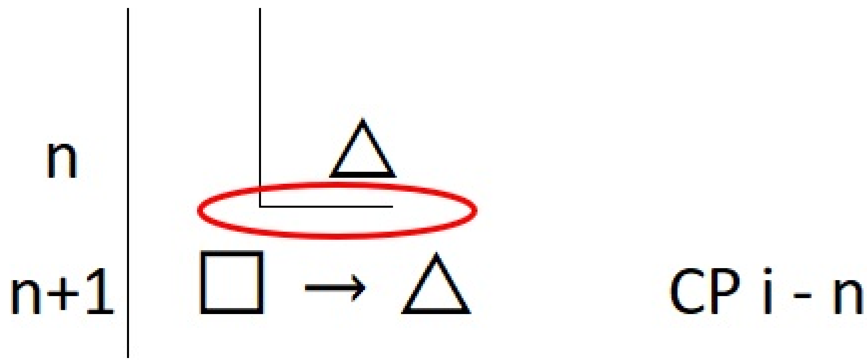
The form of the rule looks like this:



The method requires us to assume the antecedent and then get to the consequent.

The antecedent itself is NOT being proven true. We are explicitly noting that this is simply an assumption. So, our notation can be "Assume for CP." Put differently, we might say that we *suppose* that it were true. So, our notation can be "Suppose for CP." Either notation works fine. However, every other statement in the sub-proof *does need to be fully justified* in the usual ways.

Once we have arrived at a fully justified consequent, we close up the formal sub-proof. We call this "**discharging the assumption**." On paper, we'll draw a small line under the last statement in the sub-proof to show that it has closed.



At the end of the day, *we* ***must*** *discharge the assumption*. We cannot claim to have done anything if we leave the assumption open, because every statement under it has only been proven true *under* this assumption. We don't know what is true in the real world (i.e., along the primary scope line). We only know what is true *if* you help yourself to another statement which you already know may or may not be true—it's only being assumed true for the purpose of this technique.

> What *is* being shown as true in the real world is what we get when we discharge the assumption. We immediately write (along the nearest scope line) the claim □ → Δ. We know *that arrow* is true. Put differently, we know that the relationship between □ and Δ in fact holds: Δ really does follow from □.

One more note on the formal setup of CP: once you discharge an assumption, nothing contained within its scope line can be accessed or referenced directly. *Hands off!* So for example, the following would be a violation of this restriction:

1	(U v ~ D) v ~ G	
2	(A v G) · (O · ~ U)	
3	W → D	∴ (G → ~ W) · ~ U
4	G	Assume for CP
5	U v ~ D	DS 1, 4
6	O · ~ U	SIMP 2
7	~ U	SIMP 6
8	~ D	DS 5,7
9	~ W	MT 3, 8
10	G → ~ W	CP 4 - 9
11	(G → ~ W) · ~ U	CONJ 7, 10

In this example, the use of Conj attempts to reference line 7, but line 7's truth is no longer secure once we discharge the assumption after line 9.

The savvy student may already understand why this restriction is important. If something follows from an assumption, there is no real guarantee that it is true outside the scope of that assumption. So relying on that statement is sketchy at best.

> **Once an assumption has been discharged (and *it must eventually be discharged*) we no longer have access to any of those statements.**

Intuitively, CP is nothing more than a formal way to demonstrate what we mean in a conditional statement. "*If the antecedent is true, then we can be assured that the consequent is true*." We are just showing this in a step-by-step process.

You may not realize it, but you *very likely* have practiced a verbal version of this same method at many points in your life. Think of a time when one of your friends didn't believe that what you were saying was true. Like this:

> **You**: Hey, I was thinking about going to the amusement park this weekend. Do you want to come along?
>
> **Bob**: I don't know, amusement parks aren't really my thing.
>
> **You**: I get that, but this park is really special. *If we go, we're going to have an amazing time*.
>
> **Bob**: I don't know…

You: Look, let's say *we go*. Do you know what we'll see? They have some of the best rides in the country. Have you ever been on a roller coaster that has a 200-foot drop? It's such an adrenaline rush, and the view from the top is amazing!

Bob: I guess that sounds pretty cool, but I'm not really into roller coasters.

You: That's okay! You can do roller coasters, or you can do plenty of other things too. They have a fantastic water park section with some really fun water slides and a lazy river where you can just relax. Plus, there are lots of games and shows. Last time I went, they also had this incredible magic show that had everyone amazed.

Bob: I do like magic shows.

You: And the food is great too. They have this one place that sells the best funnel cakes and another spot with gourmet burgers. Trust me, if you go, it's worth going just for the food.

Bob: Hmm, I do love good food.

You: Are you kidding me! You're going to eat like a king! Plus, it's a perfect way to take a break from the usual routine. We'll spend the day outside having fun. We could even get some cool pictures and make it a day to remember. It's going to be sunny and warm this weekend, perfect weather for it.

Bob: When you put it that way, it does sound like fun. Okay, I'm in!

You: Awesome! I promise you won't regret it. *We're going to have a blast!*

You go through all the steps that connect your assumption (we go) with your final consequent (we're going to have a blast). You have now proven to Bob that it is true: *If we go to the amusement park, we're going to have a blast*. Using conditional proof is nothing more than this laid out in a formal way. Here's an example of this in symbolic form.

The completed proof looks like this:

1	$G \rightarrow [(R \cdot V) \cdot U]$	
2	$G \rightarrow [(W \cdot M) \cdot F]$	
3	$[(M \cdot F) \cdot U] \rightarrow B$	$\therefore\ G \rightarrow B$
4	G	Assume for CP
5	$(R \cdot V) \cdot U$	MP 1, 4
6	$(W \cdot M) \cdot F$	MP 2, 4
7	$W \cdot M$	SIMP 6
8	M	SIMP 7
9	F	SIMP 6
10	$M \cdot F$	CONJ 8, 9
11	U	SIMP 5
12	$(M \cdot F) \cdot U$	CONJ 10, 11
13	B	MP 3, 12
14	$G \rightarrow B$	CP 4 - 13

Note that the *conditional we want* drives the setup: we assumed *its* antecedent and have set *its* consequent as our target statement to prove.

From here, the proof proceeds as usual. We have access to the assumption we made as well as all other undischarged assumptions. In the example, this includes all the premises (remember: the *primary* scope line is never discharged). All we need to do is prove that our next subgoal (i.e., the consequent) is true.

Once completed, the annotation for CP is slightly different than most rules. We reference the entire range of the sub-proof. So we use "–" to indicate the start and end of the sub-proof.

Initially, many students ask when they should consider using CP over other ways to prove a conditional (HS or IMP). That's a fair question. There is no hard-and-fast requirement to use CP. However, the savvy student will consider the nature of CP. In using it:

a. We use a method tailor-made for proving conditionals—it's like a toaster…that thing only exists to do one thing: make toast.

So, when do you use it? Mostly, when *you want* toast

b. We get to play with another statement—we have more resources to use when we have an assumption that is temporarily true

You will find that CP greatly simplifies proofs. Sure, there are many ways to correctly finish a proof and to correctly prove a conditional to be true. However, as a general rule, unless an HS or IMP jumps out at you, you should give CP serious consideration *any time* you need to prove a conditional is true.

Nested Sub-Derivations

When we say you should consider using CP any time you need a conditional, we really mean it. Consider the following example:

1	D v G	
2	W v ~ J	
3	W → ~ G	∴ J → (S → D)
4	J	Assume for CP
5	S	Assume for CP
6	W	DS 2, 4
7	~ G	MP 3, 6
9	D	DS 1, 7
10	S → D	CP 5 - 9
11	J → (S → D)	CP 4 - 10

Our first sub-proof put us in good shape to prove our conditional. All we needed to do is prove that the consequent is true.

Q: What is the consequent?

A: It is a conditional statement (see the red emphasis)

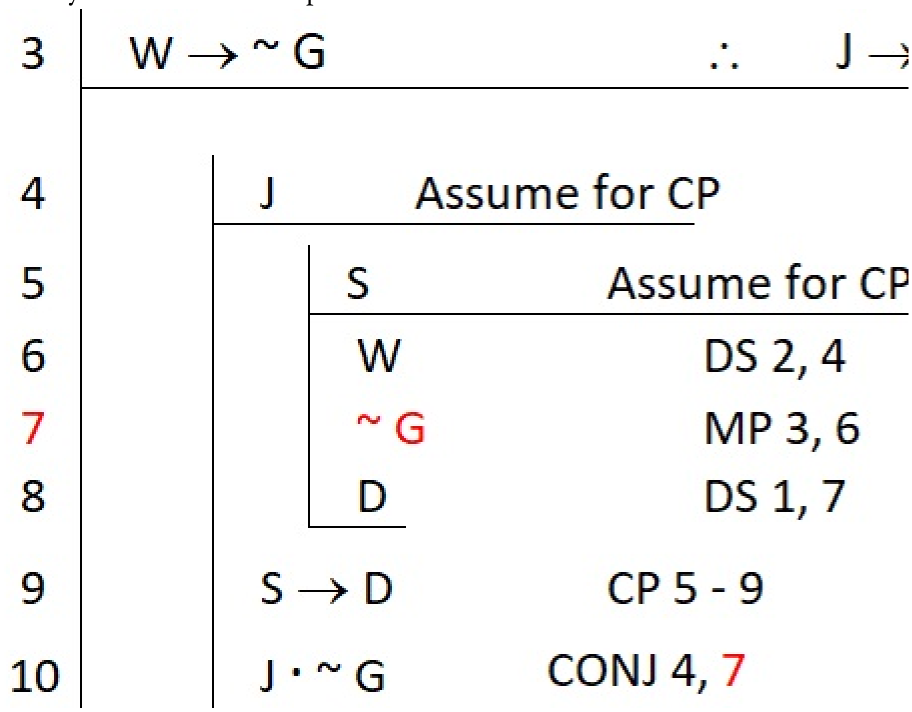
Q: Do we have any *tailor-made methods* for proving conditionals?

A: Yes, we have conditional proof…

So we just started another CP. Nothing in the CP method prohibits this. Indeed, if we needed to prove another conditional, we would be free to use it yet another time.

Of course, making these assumptions always requires us to (a) know why we are making them, and (b) to formally discharge those assumptions.

Note that our justifications for each sub-proof begin and end with the relevant assumption and the relevant subgoal. Remember that outside each closed sub-proof we have no access to the statements inside. So for example, we may have a nested sub-proof that looks like this:



Here, we might be tempted to justify line 10 by reference to line 7 (a statement that appears in the tiny sub-proof). This would not be allowed, because by the time we get to line 10 we will have already discharged the assumption under which line 7 was established.

Of course, we are free to justify line 10 with any statement that appears above and to the left of it, so long as it is not bound by a discharged assumption. For example, the reference to line 4 is perfectly fine:

1	D v G	
2	W v ~ J	
3	W → ~ G	∴ J → (S → D)
4	J	Assume for CP
5	S	Assume for CP
6	W	DS 2, 4
7	~ G	MP 3, 6
8	D	DS 1, 7
9	S → D	CP 5 - 9
10	J · ~ G	CONJ 4, 7
11		

Line 4 is not under a discharged assumption; at the time of line 10 its scope is still open. So it is fair game for justifying line 10 (so too are all the premises if they were needed). Sadly for this proof, we would still need to find a way to justify the right-hand conjunct so that it was accessible for us to justify line 10. We would need to do something like this:

1	D v G	
2	W v ~ J	
3	W → ~ G	∴ J → (J · ~ G
4	J	Assume for CP
5	S	Assume for CP
6	W	DS 2, 4
7	~ G	MP 3, 6
8	D	DS 1, 7
9	S → D	CP 5 - 8
10	W	DS 2, 4
11	~ G	MP 3, 10
12	J · ~ G	CONJ 4, 11

Here the red emphasis makes it clear that we had to repeat what we had done within the smaller sub-proof. This makes the claims on line 10 and 11 accessible to us after we had discharged the assumption that had previously closed off these claims.

If we had seen the need for these statements outside of the tiny sub-proof, we may have organized our entire strategy a bit differently to avoid the repetition. However, there is nothing inherently incorrect about the use of any of our rules in this example. The proof may look "ugly," but it is correct.

Keep such formal constraints in mind when using nested sub-proof methods.

Conditional Proof Problem Set

Conditional Proof Problem Set [PDF, 36 KB]

Indirect Proof (IP) / *Reductio ad Absurdum* (RAA)

Indirect proof is also a sub-proof method. So, it adheres to all the same general requirements and restrictions that CP follows. The assumption we make is different and the goal is different, but the general way in which it works is the same as the CP sub-proof method.

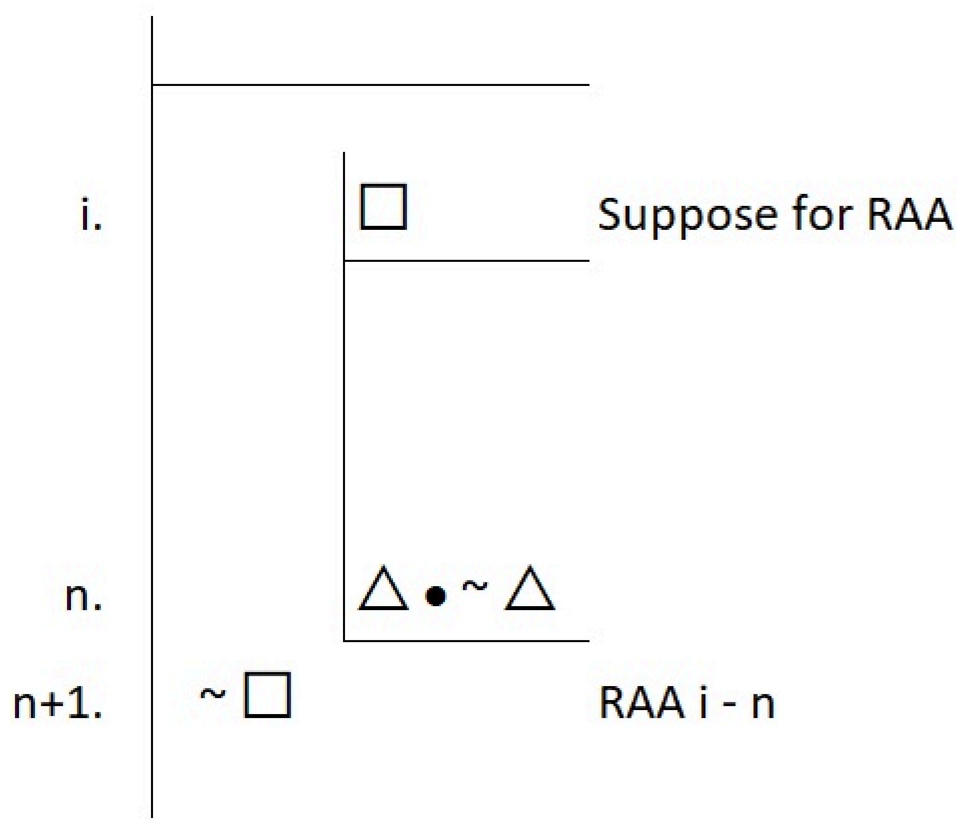
The other name for IP is *reductio ad absurdum*. This name reveals the underlying strategy of this method.

Let's say you want to diagnose a problem. You might start by making an assumption. Maybe the lights won't go on because *the bulb is bad*. Let's assume that to be true so we can test our hypothesis. If this assumption leads to some claim that simply *cannot be true*, some claim that is patently *absurd*, then there's no way our assumption can be true. We would have to deny it.

> For example, if we put that bulb in another socket and it lit up, we would know that the bulb is not bad. However, we just assumed that it was bad. So, we are forced into the claim that this bulb is both *bad* and *not bad*. That's absurd! Our assumption must be false. So we can conclude that the bulb is, in fact, not bad.

This common diagnostic process is how *reductio ad absurdum* works. In logic, claims that cannot be true are called self-contradictions. They have the form: $\Delta \cdot \sim \Delta$

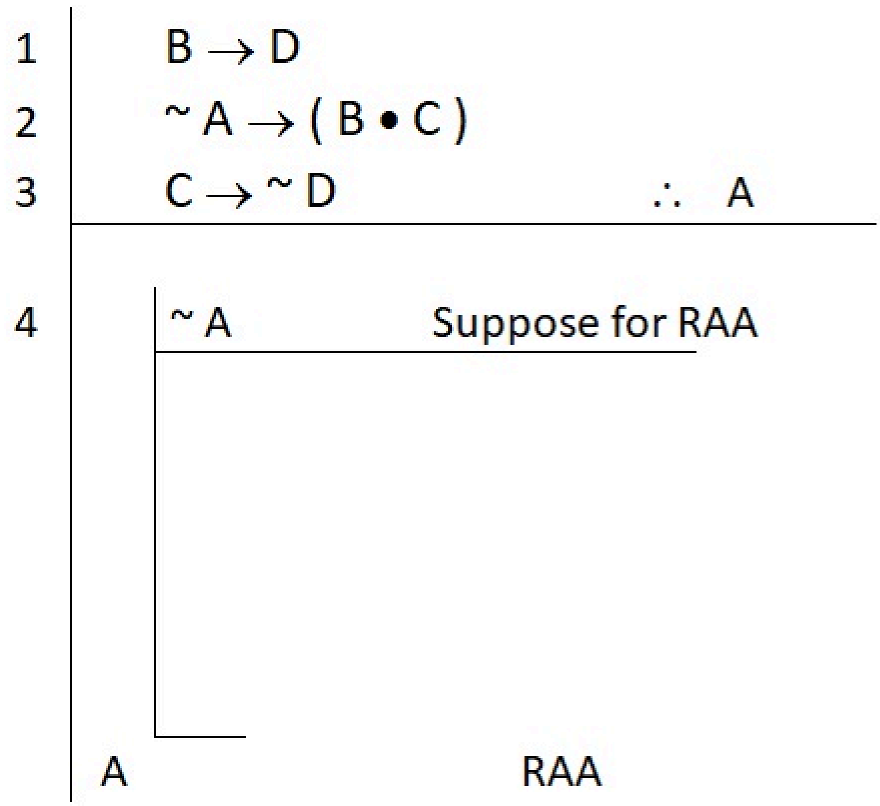
These statements are logical falsehoods. They are always false, so the principle of bivalence tells us that *the denial* of them will be true. RAA relies on the notion that an assumption that validly led to such an impossibility must also itself be false. Thus, *the negation* of that assumption must be true. The form of the method looks like this:



Strategically speaking, the setup of the RAA is simple. We know what we want. We want □. However, we *could* assume the denial of this □. We can suppose ~ □ is true. Should *that* lead us to a contradiction, then we will formally discharge our assumption and claim *that absurdity* must not stand! Thus, □ is indeed true.

Nice. But here's the rub: while *the setup* is pretty easy, *the target* is not really well known. When we look at our new subgoal, we are left with nothing much to look at except the general form of Δ · ~ Δ. We have no idea what this statement will actually be, or even if it is simply one of a number of different statements that satisfy this form. Unlike conditional proof, a RAA strategy leaves us with little guidance on what to do next. We just know we need to figure out a way to produce something absurd. *Anything absurd will do*. Anything at all, including (but certainly not limited to) □ · ~ □. We just don't know for sure…

Consider the following example:



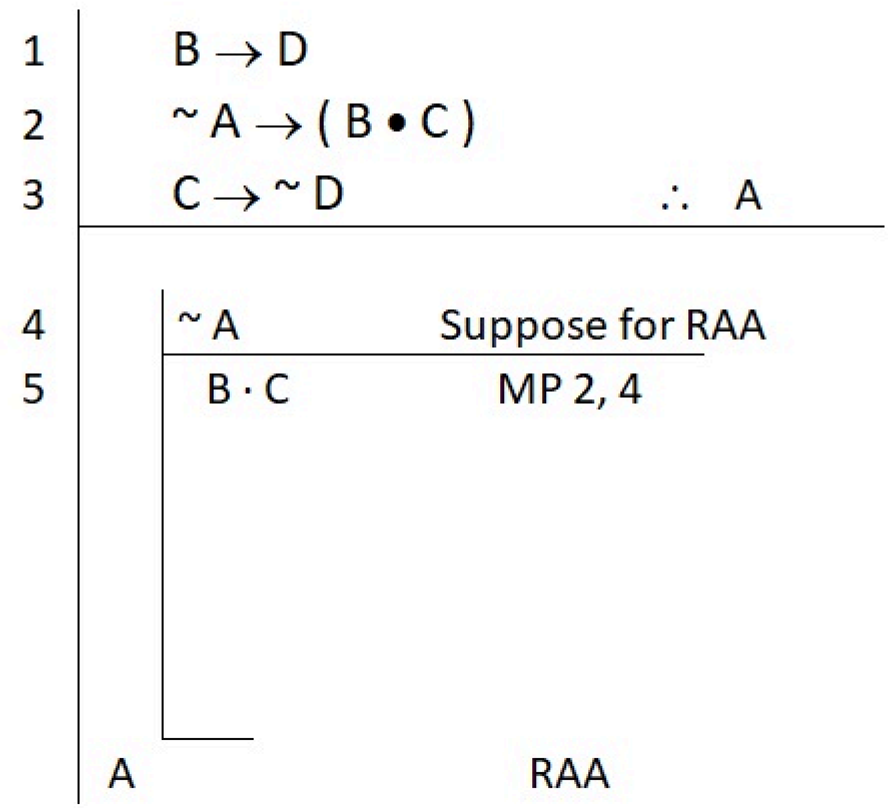
Our setup was clear enough. We wanted to prove that the atomic statement "A" is true, so we assumed the negation of it: ~ A.

From here, we need to hunt for things that look promising. RAA favors those who like to explore what the premises and their assumptions can offer.

> In using a RAA method, until you get a clear line of sight on your contradiction of choice, you will be in a mindset of "*what can I do with this?*"

In our example, we made an assumption, and we should now consider if there is anything we can do with it alongside the other statements we have access to use. Every logician will go about this in their own way—each perspective yields different opportunities. So what your friend sees as useful may differ from what you see—and both of you may be correct. Again, RAA starts off as a bit of a fishing expedition. We're just on the lookout for useful things…

We might spy this opportunity:



That assumption bought us a new statement. Can we do anything with it? Maybe this will help:

1	$B \rightarrow D$	
2	$\sim A \rightarrow (B \bullet C)$	
3	$C \rightarrow \sim D$	$\therefore A$
4	$\sim A$	Suppose for RAA
5	$B \cdot C$	MP 2, 4
6	B	SIMP 5
7	C	SIMP 5
	A	RAA

Is this going anywhere?

Don't know.

What else can we do?

Wait! What about this:

1	$B \rightarrow D$	
2	$\sim A \rightarrow (B \bullet C)$	
3	$C \rightarrow \sim D$	$\therefore$ A
4	$\sim A$	Suppose for RAA
5	$B \cdot C$	MP 2, 4
6	B	SIMP 5
7	C	SIMP 5
8	D	MP 1, 6
9	$\sim C$	MT 3, 8
	A	RAA

This is often how RAA goes. If you saw something different, great! Run it down. See if it pans out. Remember that the only requirement is to generate an absurdity. *Your absurdity may very well differ from mine.* [I suppose that's something of a logic life lesson as well.]

My finished version looks like this:

1	B → D	
2	~ A → (B • C)	
3	C → ~ D	∴ A
4	~ A	Suppose for RAA
5	B · C	MP 2, 4
6	B	SIMP 5
7	C	SIMP 5
8	D	MP 1, 6
9	~ C	MT 3, 8
10	C · ~ C	CONJ 7, 9
11	A	RAA 4 - 10

In this proof we arrived at one absurdity (C · ~ C). One is enough. Under our assumption, something intolerable has been validly derived. Thus, our assumption must be wrong. We now formally discharge our assumption and assert that its opposite is true.

Note that in many problems we will assume that a negation is true. If that assumption proves to lead to an absurdity, then the statement it negates can be asserted as true. Like this:

1. Assume ~ M
2. Get to a contradiction
3. Discharge the assumption
4. End up with M

Alternatively, we can pop out of our formally discharged assumption with a double negation. Like this:

1. Assume ~ M
2. Get to a contradiction
3. Discharge the assumption
4. End up with ~ ~ M

Either wrap-up to a RAA is fine.

Last word on RAA. This method is extremely powerful. It is also extremely risky. Because we do not have a clear sub-proof to guide us, you can get lost for a very long time in a RAA effort. You will find CP often simplifies your life and makes proofs much easier—not so with RAA.

TIP: Use RAA as a last resort *after* you feel like more conventional strategies have failed you. Unless you very quickly sense that a contradiction will be readily found, hold off on pushing the panic button.

Indirect Proof (IP) / *Reductio ad Absurdum* (RAA) Problem Set

RAA Indirect Proof Problem Set [PDF, 36 KB]

Theorems

We'll finish our section on sub-proof methods with a word on theorems in our symbolic language. A theorem is simply a statement which is logically true. Logically true statements are always true, regardless of any conditions of the world. They're weird like that. Their truth is utterly non-dependent on the facts of the world. However, this also means that it should not matter what we know about the world to prove that they are true. Put differently, you don't need *to be given any pre-existing statements* to treat as true to prove a theorem—you *don't need to rely on premises* to support the truth of a theorem. So we prove theorems with no premises. Line 1 is blank.

Of course, by now you have been practicing the reverse method of doing proofs for weeks. You don't rely on premises at all…do you…

You are a logician. Logicians rely on strategy!

To prove a theorem, you only need to understand what it is, and then generate a suitable strategy. At its deepest level, this game never depended on premises in the first place. This has always been driven by your goals. The most important logic life lesson:

> ***Goals are primary*. Where you landed in life and what you have to work with is secondary.**

So take a look at a simple theorem like this:

> ~ (J · ~ J)
>
> **Q:** What is it?
>
> **A:** It's a negation
>
> **Q:** How do you prove negations?
>
> **A:** The most straightforward way is to use RAA. Let's assume that what it negates is true; if we find a contradiction, then by RAA we will have our negation

See? *Strategy*.

Every theorem is just an exercise in pure strategy. This theorem is easy to prove. The problem starts with only a conclusion—no premises are present. Like this:

1 | ∴ ~ (J • ~ J)

We'll set up our RAA as follows:

1 | | J · ~ J   Suppose for RAA   ∴ ~ (J • ~ J)

Now the task is to find a contradiction.

Well, that shouldn't be hard…after all, we sort of assumed a contradiction to kick-start it. So all we need to do is formally demonstrate this is so *under* our assumption.

1	J · ~ J	Suppose for RAA	∴ ~ (J • ~ J)
2	J	SIMP 1	
3	~ J	SIMP 1	
4	J · ~ J	CONJ 2, 3	
5	~ (J • ~ J)	RAA 1-4	

This was always going to be an easy proof. The theorem was denying an absurd statement in the first place. So, no wonder RAA successfully proved *the denial* of absurdity *is true*.

Of course, not every theorem is a negation. Consider the following simple theorem:

(E · M) → (~ G v M)

Q: What is it?

A: It's a conditional statement

Q: Do we have any tailor-made methods for proving conditionals?

A: Yes, we have conditional proof…

The setup for a conditional proof does not require premises. The setup only requires knowledge of *what you want*. So we set it up as follows:

1	E • M	Suppose for CP	∴ (E • M) → (~ G v M)
	~ G v M		
	(E • M) → (~ G v M)		

From here we now know that we need to show that the consequent (~ G v M) is true. This is easy:

1	E • M	Suppose for CP	∴ (E • M) → (~ G v M)
2	M	SIMP 1	
3	~ G v M	ADD 2	
4	(E • M) → (~ G v M)	CP 1 - 3	

After a few of these, the savvy student realizes that proving theorems is actually quite easy. They will require the use of one of the sub-proof methods to get started. From there, the proofs follow the usual strategies.

Keep in mind that sub-proof methods have no restrictions on when they are used. In other words, similar to what we saw in nested conditional proofs, you may use *a combination* of CP and RAA within one another to get the job done.

Theorems Problem Set

Theorems Problem Set [PDF, 26 KB]