Inductive Reasoning

PHIL 145 Textbook Supplement

Chapter 1
Inductive Arguments: Strength, Cogency, Completeness

When we reason deductively, we intend that our premises will *guarantee* the truth of our conclusion. *Valid* deductive arguments are deductive arguments that do this successfully: given that their premises are true, it is impossible for their conclusion to be false.

Not all good reasoning is deductive. It is possible for an argument to give us good reasons to believe something, but not to *prove* that thing. Such reasoning is very important. Life requires that we make decisions about uncertain matters. Paradoxically, it is often when we are reasoning about the things that matter the very most to us that deductive certainty is the hardest to come by. Doctors often have to make educated guesses about what illness a patient has; if they wait for conclusive proof of the diagnosis, the window of opportunity for treating the illness will have closed. Or, again: jurors in a courtroom have to decide the guilt or innocence of the defendant, though the evidence is never going to add up (in a timely fashion, anyway) to a deductive proof one way or the other. Juries are told to treat defendants as innocent unless the evidence points to their guilt "beyond reasonable doubt." That's a high bar, but it's not deductive proof, which would require that the evidence leave room for no doubt whatsoever.

Inductive reasoning is thus also known as "ampliative" reasoning. It does not just take the information in the premises and manipulate it using formal rules (as is the case with deductive reasoning). Rather, it permits the information in the premises to be 'amplified' into a conclusion that is not actually guaranteed by the premises. This doesn't mean, of course, that inductive arguments permit us to jump to whatever conclusions we wish. But the difference between 'jumping to conclusions' and engaging in careful ampliative reasoning isn't always obvious, and people mistake the former for the latter all the time. The many successes of modern science in telling us about the natural world show us how powerful inductive reasoning can be—and how meticulous we must be to engage in it well. Because inductive reasoning lets us 'go beyond' the information in the premises, it is inherently risky—and potentially very powerful.

To begin to see how inductive arguments can go well or badly, let's think through a mundane example. It is an October morning in northwest Indiana. You're about to walk out your front door, and you're wondering whether to bring a jacket: if it's going to be warm later, you don't want to be burdened with the extra garment, but if it's going to be cool, you don't want to be underdressed. You step outside. The air temperature is comfortable, if not particularly warm, and the sun is shining. You quickly infer that you'll be warm enough without your jacket.

We can reconstruct your reasoning as an argument in standard form (enumerated premises followed by a conclusion) as follows:

P1. The vast majority of comfortable, sunny fall mornings are mornings I do not need to bring my jacket with me in order to be warm enough.
P2. This morning is a comfortable, sunny fall morning.
C. So, This morning is a morning I do not need to bring my jacket with me in order to be warm enough.

(This is a substitution-instance of a *proportional syllogism*, which is the first of five types of inductive argument we will cover in this unit.) Is your reasoning valid? No. It is easy to construct an informal counterexample to demonstrate this:

> We can grant that the vast majority of comfortable, sunny fall mornings are not mornings I need to bring my jacket with me in order to be warm enough and that this morning is a comfortable, sunny fall morning. It doesn't follow that this morning is not a morning I need to bring my jacket with me. After all, it could be the case that a cold front moves in today, and the afternoon turns out to be quite chilly.

But just because your reasoning isn't valid doesn't mean it's no good: though the premises don't *guarantee* that your conclusion is true, they do make it very likely. And that is to say that the argument is **inductively strong**: *given that the premises are true, it is improbable that the conclusion is false.*

Just *how* probable does the conclusion have to be, given the premises, for the argument to be strong? At the very least, (a) the premises need to *raise* the probability of the conclusion, and (b) the probability of the conclusion, given the premises, needs to be greater than 50%. If either of these conditions fails to hold, the argument is weak. Suppose, for example, that I make the following argument:

P1. You just got a phone call from one of your parents.
C. So, you got a phone call from your father.

Without any further information (about the frequency with which each of your parents calls you, say), it's only 50% likely that your father is calling you rather than that your mother is calling you. So, the argument is clearly weak.

Now, what about an inductive argument whose premises make the conclusion 51% likely? Would the argument then be *strong*? Well, *not very*. In other words, inductive strength comes in degrees. We can ask, of an inductive argument, not just *whether* it's strong but *how* strong it is. (Notice: it makes no sense to ask, of a deductive argument, how valid it is. A deductive argument is either valid or invalid, and that's the end of it.)

Some inductive arguments allow us to actually *calculate* the degree of inductive strength. For example, suppose you're playing Texas Hold 'Em poker with three other players. Each player has been dealt two cards. You have an Ace and a Jack; you gather from another player's betting behavior that she has an Ace, too. The dealer has dealt "The Flop" and "The Turn", so now there are four cards face-up on the table, and two of these cards are Jacks. One more card remains to be dealt ("The River"). If it's an Ace, you'll have a full house. Should you bet on getting that full house? No way. There are only 2 Aces left in the deck, and 40 more cards. That means there's a 38/40 chance that you won't get your full house. You infer that you won't get an Ace. And your inference is quite strong: 95%.

Now, in this case, it was pretty easy to calculate the probability of the conclusion (that you wouldn't get an Ace). All you had to do was subtract the number of dealt cards (12) from the total number of cards in a deck (52); figure out how many cards there are in the remaining deck that *aren't* Aces (38); and then divide the number of non-Aces in the deck (38) by the number of remaining cards in the deck (40). But probability calculations are rarely this simple. For example, if you were asked to calculate the probability that there are two students in this class who share a birthday, would you immediately know how to do it?

Mathematicians have developed a system of rules that tell us how to approach puzzles like this. The system is called 'The Probability Calculus'. The Probability Calculus is to inductive reasoning what the natural deduction proof system (the 18 rules and the way they are applied) is to deductive reasoning. It is the 'logic' of inductive reasoning.

The Probability Calculus is very powerful as a way of determining the degree of strength of many inductive inferences. It has a crucial limitation, however: it only applies to inferences whose premises have numerical content of some sort. Consider the following case: you're weighing evidence as a juror in a murder trial. On the one hand, the finger-prints of the accused were found on the gun that was used to kill the victim, and the accused was very open about his intense hatred for the victim. On the other hand, he swears he didn't do it, he has no history of violence, and his wife says he was at home at the time of the murder. What is the probability—in terms of a number between 0 and 100—that he is the murderer, given these bits of information? In order to answer this question, we would first need to know the probability that he is the murderer given *each* of the bits of evidence individually, and we would need an equation that generated a single number out of all of these individual probabilities. (There is a fascinating sub-discipline of philosophy, known as "Formal Epistemology," that tries to do this. Its central tool is the probability-rule known as "Bayes' Theorem.") However useful it may be to model inductive reasoning on the mathematics of probability—and it is certainly a matter of debate just how useful it is—we will not be exploring this strategy. Instead, we will learn how to recognize and reconstruct inductive arguments of five specific types, and we will discuss criteria, unique to each type, that can be used to assess its degree of strength.

So far we have been discussing the 'logical analysis' of inductive arguments—that is, how we evaluate the *quality of the inference* they make from premises to conclusion. And we said that an inductive argument is *inductively strong* when the premises make the conclusion probable (where this probability comes in varying degrees). But we can also subject inductive arguments to *truth-value analysis* by asking whether the premises are true. Just as we have a term for deductive arguments that are valid *and* have true premises (they are "sound"), so we also have a term for inductive arguments that are strong and have true premises: they are **cogent.** Uncogent arguments are inductive arguments that are either weak or that have one more or more false premises (or both).

There is a way for an inductive argument to be cogent—strong, with true premises—yet remain totally unpersuasive. To see how, consider the following argument:

P1. The answer that Tanya Shechtmann, the president of the MIT physics club, gave to problem #4 on the Introduction to Physics homework is that acceleration due to gravity on the surface of the moon is approximately 2.7 m/s^2.

C. So, acceleration due to gravity on the surface of the moon is approximately 2.7 m/s^2.

This argument is strong, and—if the premise is true—cogent. (It is an instance of the type of inductive argument known as "Argument from Testimony"—an inference from somebody's having made some assertion, to that assertion's being true. Such arguments are never deductive, but they can be quite strong when the person is question is an authority on the subject-matter, as Tanya appears to be.)

But let's say we hear from a friend that Tanya was taking Tequila shots while she was doing her homework. Should we embrace the conclusion? No. True, the premise does make the conclusion probable (given common knowledge about MIT and so forth). But this premise *in tandem with our new bit of information* does *not* make the conclusion probable. Our new bit of information 'blunts' the probabilifying force of the premise. We might say of the argument: "It's great, so far as it goes, but it is crucially incomplete." Note that there are lots of other things we could learn about Tanya that are potentially relevant—she has an extremely high alcohol tolerance; she became president of the physics club by bribing someone; she thinks it's fun sometimes to give wrong on answers on homework assignments; etc.

This reveals another important contrast between deductive and inductive arguments. If the premises of an argument make the conclusion 100% likely—that is, if it's a valid, deductive argument—there is no new information that you could add that would make the conclusion less than 100% likely. But if the premises make the conclusion any less than 100% likely—that is, if it's an inductive argument—new information ('new' in the sense that it is not packed into or implied by the premises) can be combined with the premises that lower the likelihood of the conclusion. This means that deductive arguments can be fully evaluated simply by working with the materials included in the premises and the conclusion: you can 'bracket' everything else you know, and just use logic. But inductive arguments cannot be 'bracketed' in this way. Anything you know that is relevant to whether and to what degree the premises support the conclusion is 'on the table' when evaluating an inductive argument.

New information, then, can blunt the probabilifying force of the premises of an inductive argument, even a strong one. When we are in possession of such information, we will judge the argument **incomplete** and find it correspondingly less persuasive than it otherwise would be. In sum: to be fully persuasive, an inductive argument must be not only cogent but **complete**: there must be no new information in our possession that significantly blunts the probabilifying force of the premises.[1]

Chapters 2-7 cover five types of inductive argument. They are:

[1] Philosophers have a term for new information that has this effect on an argument: it serves as an "undercutting defeater." Undercutting defeaters contrast with "overriding defeaters," which are bits of new information that tell against the truth of the conclusion. For example, if our textbook tells us that acceleration due to gravity on the surface of the moon is approximately 3.7 m/s^2, this new evidence "overrides" the evidence of Tanya's answer. Undercutting defeaters aren't like this. Learning that Tanya was drunk while doing her homework doesn't tell us that the conclusion of the argument is *false*. It just undercuts the support that we previously had for that conclusion.

Proportional syllogism: an inference from a claim about an entire category to a conclusion about a member of that category

Argument from analogy: an inference from a claim about a thing or category to a conclusion about a similar thing or category

Inductive generalization: an inference from a claim about a sample of a category to a conclusion about the whole category

Causal Inference: an inference from observed correlations among things, categories, or events to a conclusion about a causal relationship between those things, categories, or events

Inference to the best explanation: an inference from an observed phenomenon to the best explanation of that phenomenon

Chapter 8 then considers reasoning about a particular subject matter: *morality*. Moral arguments come in deductive and inductive varieties, and we will discuss some of both.

Chapter 2
Proportional Syllogism

2.1 Reconstructing Proportional Syllogisms

The following is a valid categorical syllogism:

P1. All Estonians are good singers.
P2. Stavra is an Estonian.
C. So, Stavra is a good singer.

(To demonstrate its validity, all we need to do is paraphrase the second premise as "All people identical to Stavra are Estonians", and the conclusion as "All people identical to Stavra are good singers"—and then diagram it using a Venn diagram.)

Because the argument is valid, the conclusion follows necessarily from the premises: we can be *certain* about Stavra's vocal skills, assuming the premises are true. But premise 1 is not really plausible. It says that *all* Estonians—100% of them!—are good singers. Even if bad Estonian singers are rare, there are probably a few of them. Let's say that the Estonian Bureau of Choral Suitability keeps records on such matters, and according to its most recent numbers, 95% of Estonians are good singers. We can alter our argument accordingly:

P1. 95% of Estonians are good singers.
P2. Stavra is an Estonian.
C. So, Stavra is a good singer.

We can no longer be certain of the conclusion, given the premises, so it's no longer a valid argument. But we can still be *pretty* sure of the premises. After all, we know that a really high percentage of Estonians are good singers, so it's a good bet that Stavra (being Estonian) is a good singer. That is to say: our modified argument is a strong inductive argument. Given the premises, the conclusion is highly probable. It is, moreover, a strong inductive argument of a particular form—a *proportional syllogism*. Construed very generally, a proportional syllogism is an inference from a claim about an entire category to a conclusion about a member of that category. Proportional syllogisms have the following form:

P1. n% of As are F.
P2. x is an A.
C. So, x is F.

The first premise says, of some **population** A that a certain **proportion** n of that population has some **target feature** (F). (In our example, the population is *Estonians*, the target feature is *being a good singer*, and the proportion is *95%*). The second premise says that some **selection** (x) is a member of the population in question. (In our example, the selection is *Stavra*. She is, at it were, "selected" from among all the people who comprise the population *Estonians*). Finally, the conclusion states that the selection has the target feature. The target feature is so-called because it is the target of our inquiry. We want to know, of the selection, whether it has *this* feature.

Proportional syllogisms are very common, though they are hardly ever put in standard form. Recognizing and reconstructing proportional syllogisms is not difficult, so long as we keep in mind (a) the general idea—an inference from a claim about an entire category to a conclusion about a member of that category—and (b) the precise details specified by the form. Here is an example:

> 80% of patients who undergo a kidney transplant require dialysis for at least six months afterward. You'll be undergoing such a transplant, so you'll be needing to continue kidney dialysis for six months after your operation.

Though this bit of text is not in standard form, it is clearly an inference from a claim about a general category (kidney transplant patients) to a member of that category (you). Reconstructing it as a proportional syllogism is a matter of identifying the items that replace the variables in the form. We have already identified two:

> A: Kidney transplant patients
> x: you

All that's left are F and n. n is obviously 80. Identifying F is also pretty easy. We simply need to figure out what the conclusion of the argument is, since—as the form makes clear—the argument will mention both the selection and the target feature, ascribing the latter to the former ("x is F"). (Of course, F also shows up in the first premise.) So, we have:

> F: requiring kidney dialysis for at least six months

Now we simply build back to a full substitution-instance of the form, using the items we have identified:

P1. 80% of kidney transplant patients require kidney dialysis for at least six months.
P2. You are a kidney transplant patient.
C. You require kidney dialysis for at least six months.

This is very close to the original. Note that the language has been standardized across premises. In the original, the first premise included the verb 'require' whereas the conclusion uses 'be needing'. Such standardization is an important component of argument reconstruction, even if the resulting language sounds a little boring or stilted. Second, the language has been streamlined: 'afterward' in the first premise and 'after your operation' in the conclusion have been omitted. We could have included them in our reconstruction, but we don't have to. Third, note that the conclusion lacks an 'is', or even the grammatically correct 'are' that would have to follow the subject 'You'. We *could* have included a copula, but the result would have been awkward: "You are requiring kidney dialysis for at least six months." But because we're not going to be *diagramming* proportional syllogisms, non-standard verbs can be used in place of a copula. The bottom line is that we shouldn't make any additions or subtractions that change the meaning. If we're following that rule, some latitude is permitted.

As common as proportional syllogisms are, it is not common for the speaker to make each statement in the argument fully explicit. More commonly, one or more of the statements in the argument will be implied. (Logicians call arguments with implicit premises or conclusions 'enthymemes'.) But so long as we recognize a bit of text as expressing the general idea of a proportional syllogism—making an inference from what is known about an entire category to a conclusion about a member of that category—and so long as we make use of the form, we can readily fill in the missing details. Here is an example:

> Georgia probably doesn't use ranked-choice voting, since only 6 of the 50 states do so.

'Since' is a premise-indicator, so the first statement is our conclusion. It gives us the selection and target feature easily:

> x: Georgia
> F: not using ranked-choice voting

What's the population? This, too, is not difficult. Remember that the selection is always a member of the population-category. What category is Georgia a member of?

> A: The 50 states

What about n, the proportion? In order to make sense in the first premise, it needs to be the proportion of the 50 states that have the target feature, which is *not* using ranked-choice voting. We're told that 6 of 50 states *do* use ranked choice voting, so we first need to subtract 6 from 50 to calculate the number of states that *don't*, i.e., 44, and then divide by 50:

> n: 88

Now we can reconstruct the original, enthymematic text as a standard-form proportional syllogism:

P1. 88% of the 50 states do not use ranked-choice voting.
P2. Georgia is one of the 50 states.
C. So, Georgia does not use ranked-choice voting.

(Notice that 'probably' showed up in the original text, but has dropped out of our reconstruction. This is because 'probably' is an *inductive-inference indicator*, rather than part of the argument proper.)

2.2 Evaluating Proportional Syllogisms

Now that we have learned to recognize and reconstruct proportional syllogisms, how do we assess them for inductive strength? The answer couldn't be simpler: *n* tells us exactly how strong the argument is. If n is less than or equal to 50, the argument is weak; if it is greater than 50, the argument is strong, and n is the measure of its strength. This is because the probability of the conclusion given the premises is equivalent to the proportion of the population known to have the target feature. The more widespread the target feature in the population, the greater our degree of confidence that the selection has that feature.

Now, speakers don't always have precise statistical information about the population in question; sometimes they have to get by with saying that 'most' members of the population have the feature, or 'nearly all', 'virtually every', or similar, and we have to use these vague expressions in place of 'n' when doing our reconstruction. When this happens, we have to be equally vague in our assessment of the inductive strength of the argument; we'll say that the argument is 'somewhat strong' or 'very strong', as best we can tell. (Note that 'many' and 'lots of' and so forth are not equivalent to 'most' and 'nearly all'. These expressions can be used even when the proportion of the population that has the target feature is less than 50%. For example: "many professors are wealthy" can be true even if most professors are not.)

<u>Chapter 2 Homework Exercises</u>

Instructions: Each of the following contains a proportional syllogism. Reconstruct it as a substitution-instance of the form discussed in this chapter and then specify whether it is strong or weak. (Note that a premise or conclusion may be implied but unstated.)

Example

Encouraged by a producer she met while working as a paralegal in New York City, Janae Stewart moved to Hollywood in 2007. She hoped to make it as a professional Hollywood screenwriter. The truth is that working as a screenwriter in Hollywood is a largely futile endeavor: 95% of scripts never even get looked at by a production company, let alone get put into production. But producing a new script can take years. It is very unlikely that Stewart's scripts will ever make it to the screen. In 2012, Stewart resolved that if none of her work had gotten picked up by the end of the year, she would return to law. Four years later, she continues to live and write in Hollywood. 'This year really will be my last,' she says with a smile and a sigh.

Answer:
P1. 95% of scripts written by a Hollywood screenwriter are unproduced.

P2. Stewart's scripts are scripts written by a Hollywood screenwriter.
C. Stewart's scripts are unproduced.

Strong

1. Eleanor Rigby is a student at Valparaiso University. So, Eleanor must be a basketball fan, since 8 out of 10 VU students are basketball fans.

2. I dislike 19 of the 22 operas Mozart wrote. So, I'll probably dislike the opera we're going to tonight, since it's one of Mozart's.

3. James Donahue's injury is a gun-shot wound. Over 99% of gun-shot wounds are received in a context other than a mass shooting, so his injury was almost certainly the result of some other type of shooting.

4. On average, there is snow on the ground in Chicago only two out of every three Christmases. There's a good chance, then, that we won't have to deal with snow in Chicago on December 25 this year.

5. I took a peek at the passenger manifest. Of the 47 passengers aboard the train, 21 are married. Alas, the attractive man sitting across from me is probably married, then.

6. I will probably die in some manner other than a car accident, since only about 1 in 10,000 Americans dies in a car accident.

7. Out of the 350,000 Subaru Outbacks manufactured between 2000 and 2004, roughly 200,000 need a head gasket replaced when they reach 120,000 miles. My Subaru Outback was manufactured during that window. So, I will need to get my vehicle's head gasket replaced when it reaches 120,00 miles.

8. Irene and Tobey are millennials who married last year. 75% of millennial couples are still married five years after their wedding. So, Irene and Tobey have a good chance of making it to their fifth wedding anniversary.

Chapter 3
Argument from Analogy

3.1 Reconstructing Arguments from Analogy

Let's say you find yourself, for the first time, at O'Malley's Irish Pub with a couple of friends. Scanning the menu, you notice that the restaurant serves steak-and-ale pie. You are a big fan of steak-and-ale pie when it is done right, but you also know that beyond the shores of the British Isles it is *not* always done right. How could you determine, of *this* restaurant's steak-and-ale pie, that it is of high culinary quality?

Notice that you are probably not in a position to construct a proportional syllogism. In order to do so, you would have to know the statistical frequency with which the target feature—in this case, being of high culinary quality—occurs in some population of which the pie in question is a member. For example, you would need to know that

> 80% of steak-and-ale pies made in the U.S.A. are of high culinary quality, *or*
> 80% of dishes served at O'Malley's are of high culinary quality

or something similar. But, of course, you aren't in possession of such information. Now, there are *methods* for acquiring such information. In the next chapter, we will be discussing them. But these methods take a great deal of time and care and you're hungry. What do you do?

If you're like me, you ask your friends at the table whether any of them have had O'Malley's food before. One tells you that the lamb stew is extraordinary. The other tells you that they make the best bread pudding in the Region. You're persuaded: you feel ready to take a chance on the steak-and-ale pie.

Let's think about why this made sense. What you have done is found items that have the target feature (they're of high culinary quality), and that are also *similar to* the item in question (the steak-and-ale pie) in relevant ways: in particular, they are traditional Irish foods that are served at O'Malley's. Then you *draw an analogy* between the foods you know to be good and the dish you're trying to learn about. We can reconstruct the inference you have made as follows:

P1. The lamb stew and the bread pudding are traditional Irish foods made by O'Malley's and are of high culinary quality.
P2. The steak-and-ale pie is a traditional Irish food made by O'Malley's.
C. So, the steak-and-ale pie is of high culinary quality.

There's no guarantee, of course, that the steak-and-ale pie is of high quality (which is just to say that it's not a deductive inference). Maybe there is some crucial difference between the pie and the two other dishes (it's a trickier dish to make than the other two; a different chef prepares them, etc.). But you aren't aware of any of these differences, and you *are* aware of a couple of crucial similarities: being traditional Irish foods; being made by O'Malley's. So the inference looks fairly strong. In particular, it looks to be a strong *argument from analogy*. Construed very generally, an argument from analogy is an inference from a claim about a thing (or category) to a conclusion about a similar thing (or category). Arguments from analogy have the following form:

P1. x1, x2, x3, etc.… are A, B, C, etc., and F.
P2. y is A, B, C, etc.
C. So, y is F.

At the heart of an argument from analogy is a comparison between one or more items you already know about and an item you are trying to learn about. The known-about items are called "**primary analogues**" (in our example, the lamb stew and the bread pudding) and the to-be-learned-about item is called the "**secondary analogue**". (Calling them 'analogues' says nothing more than they are things between which an analogy is being drawn.) Let's go through the form in reverse. The conclusion says that the secondary analogue has the **target feature** F, the feature that is the target of our inquiry (in our example, being of high culinary quality). The second premise says, of the secondary analogue, that it has certain features A, B, C, etc., called "**base features**"—the features that the primary and secondary analogues *have in common*, and on the basis of which a legitimate analogy can be drawn between them (in our example, being traditional Irish foods and being made by O'Malley's). The first premise says, of the primary analogues, that they have the base features, and that they have the target feature.

Arguments from analogy are very common: we're employing them whenever we make comparisons between familiar things and unfamiliar things in order to learn more about those unfamiliar things. One domain where arguments from analogy reign supreme is animal testing. Suppose, for example, that we want to know whether a new drug, Zypheralax, is likely to be effective as a cholesterol-reducer, so we test it on mice. Here's the reasoning we use:

> Mice are mammals, have a mammalian circulatory system, respond readily to high blood pressure drugs, and experience a reduction in blood cholesterol when given Zypheralax. Humans are mammals, have a mammalian circulatory system, and respond readily to high blood pressure drugs. So, humans will also experience a reduction in blood cholesterol when given Zypheralax.

Let's identify each of the items in the form. Primary and secondary analogues are the items that are getting compared. Note that in this case, there's only one primary analogue. (If you're ever having trouble deciding which analogue is primary and which is secondary, look to the conclusion: only the secondary analogue is mentioned there.)

x: mice
y: humans

The target-feature is the feature that, in the conclusion, is attributed to the secondary analogue (humans):

F: experiencing a reduction in blood cholesterol when given Zypheralax

Finally, the base-features are the features in common between the two analogues (mice and humans):

> A, B, C, etc.: being mammals, having a mammalian circularity system, responding to high blood pressure drugs

Here, then, is our resulting reconstruction:

P1. Mice are mammals, have a mammalian circulatory system, respond to high blood pressure drugs, and experience a reduction in blood cholesterol when given
P2. Zypheralax.Humans are mammals, have a mammalian circulatory system, and respond to high blood pressure drugs.
C. So, humans experience a reduction in blood cholesterol when given Zypheralax.

This is, of course, very similar to the original text, though now in standard form (and wearing its structure 'on its sleeve,' as it were). One thing to notice is that the primary and secondary analogues are not individual items, but rather whole categories. This is a perfectly legitimate use of argument from analogy.

It's worth pointing out that base features are hardly ever spelled out this explicitly, let alone mentioned twice (once in connection with mice, and again in connection with humans). Sometimes the speaker leaves it to the hearer to intuit what the base-features are—because the similarities are too obvious to mention, perhaps. But sometimes the speaker doesn't mention the base features because they are actually pretty minimal, and suppressing them makes the argument seem stronger than it is—a rhetorical trick that loses its effect once the argument is reconstructed.

3.2 Evaluating Arguments from Analogy

An inductively weak argument from analogy is called a "false analogy." There is no simple way to determine whether an analogy is a 'true' one, nor is there a way to quantitatively measure the degree of strength of an argument from analogy. So we have no way to answer the question, "how strong is this argument from analogy?" in a precise and numerical way. Nevertheless, we can identify dimensions along which arguments from analogy can be stronger or weaker: (1) the quantity and relevance of the base features, and (2) the number and diversity of primary analogues.

(1) *The quantity and relevance of the base features.* Let's work through a different example. Suppose you decide to take a Macroeconomics course. (Maybe you find the topic fascinating and don't want to miss the chance to learn from an expert; maybe it satisfies a curricular requirement.) The section that best fits into your schedule is taught by Prof. Klimp. You have never taken a course from Prof. Klimp before, but you do know three fellow students who have taken Prof. Klimp's course: Erin, Doug, and Sylvester. Erin, who like you is majoring in Finance, liked Prof. Klimp's Macroeconomics course. You infer that you'll like it, too. We can construct your inferences as follows:

P1. Erin is a Finance major and liked Prof. Klimp's Macroeconomics course.
P2. I am a Finance major.
C. So, I will like Prof. Klimp's Macroeconomics course.

How strong is this argument? Intuitively, not that strong. (I'll call it the 'Weak Version'). Why? Well, you're basing your inference on *one* similarity between you and Erin. One way to improve upon the Weak Version would be to add more base features. Maybe you and Erin have the same opinion of two other courses that you have both taken. Maybe you both listen to the same Economics podcasts. These similarities between you and Erin certainly suggest that you will like an Economics course that she liked. Here, then, is the First Stronger Version of your inference:

P1. Erin is a finance major, liked courses X and Y, disliked course Z, listens to Podcasts A and B, and liked Prof. Klimp's Macroeconomics course.
P2. I am a finance major, liked courses X and Y, disliked course Z, and listen to Podcasts A and B
C. So, I will like Prof. Klimp's Macroeconomics course.

There are probably lots of other similarities you have with Erin. Let's say you both are fans of the Detroit Tigers; you have the same shoe size; you both write poetry. Should we add these similarities to the list?

We could, but doing so probably wouldn't help matters. What matters to the strength of the argument is that the primarily and secondary analogues have base features in common *that are predictors of the target feature*. That is, the base features should be the sorts of features whose presence make the presence of the target feature more likely. In short, base features need to be "relevant" to the target feature in order to make the argument strong. Neither baseball franchise fandom, nor shoe size, nor poetry authorship seem relevant to whether one likes Macroeconomics, at least not in any obvious way. Now, maybe they're relevant in some non-obvious, indirect way. (Maybe people who grew up near Detroit tend to develop an affinity for both the Tigers and for economic theory.) But unless we have reason to think so, we have to treat these features as adding nothing to the strength of the argument.

All other things being equal, a greater quantity of base features *does* strengthen an argument from analogy. But the boost might be very small if the features are, as best we can tell, irrelevant. A few, clearly relevant base features is better than many, irrelevant base features.

(2) *The number and diversity of primary analogues*. There is a different way to strengthen the Weak Version. Instead of (or, in addition to) adding base features, you could add more primary analogues: you could ask Doug and Sylvester what they thought of Prof. Klimp's Macroeconomics course. And let's say both of them enjoyed the course. What emerges is stronger than the Weak Version, which I'll call the 'Second Stronger Version':

P1. Erin, Doug, and Sylvester are Finance majors and liked Prof. Klimp's Macroeconomics course.

P2. I am a finance major.
C. So, I will like Prof. Klimp's Macroeconomics course.

This version is stronger than the Weak Version. So, in general, the greater the number of primary analogues, the stronger the argument. But quantity can be misleading here. Let's suppose that Erin, Doug, and Sylvester are all minoring in Economics. Maybe this is this is really what explains why they liked the course. But you're *not* an Economics minor. You do share some base features with Erin, Doug, and Sylvester, but those features aren't as predictive of the target feature as is a feature you *don't* share with them. Something has gone wrong.

What has gone wrong is that your primary analogues are too similar to one another. As a consequence, adding primary analogues hasn't done much to strengthen the argument. What *would* strengthen the argument is if you found three Finance students who liked Prof. Klimp's course, but who otherwise are very different from one another. Doing so would help guarantee that the base feature (in this case, being a finance major), and not some commonality among the primary analogues, predicts the presence of the target feature (liking the course).

In sum: all other things being equal, a greater quantity of primary analogues strengthens an argument from analogy. But the boost might be very small if the primary analogues are carbon-copies of one another. Better is for them to be quite different in all other respects besides the base features. Now, even if an argument from analogy is strong in virtue of having the two features we have discussed—it invokes multiple, diverse primary analogues that all bear multiple, relevant similarities to the secondary analogue—there remains a crucial question to ask about whether it is *complete* (i.e., whether there is any new information that significantly blunts the probabilifying force of the premises). In particular, we need to ask whether the *differences* between the primary and secondary analogues outweigh the *similarities*. We'll treat this *completeness* consideration as a third dimension along which we need to evaluate arguments from analogy, as follows.

(3) *The extent of the differences*. Suppose that Erin, Doug, and Sylvester all studied Calculus in high school and enjoyed doing so. But you didn't make it any further than Geometry and you hated every second of it. Clearly, once we take this new information into account, the Second Stronger Version is much less persuasive:

P1. Erin, Doug, and Sylvester are Finance majors [and excel at math] and liked Prof. Klimp's Macroeconomics course.
P2. I am a Finance major [but struggle at math].
C. So, I will like Prof. Klimp's Macroeconomics course.

(The new information is in brackets because it isn't part of the argument; it is new information that renders the argument less persuasive.) The point is not just that there *is* a difference between you and the three other students. There *always* are differences between primary and secondary analogues; that's what makes arguments from analogy inductive. The point is that the difference in question seems highly relevant to the target feature. It thus counts as new information that blunts the probabilifying force of the premises.

Evaluating an argument from analogy for *completeness* is thus as important as evaluating it for *strength*. Really, the two tasks go hand-in-hand: to evaluate an argument from analogy, one must do the work of comparing and contrasting the primary and secondary analogues. One must ask: are the similarities between the primary and secondary analogues extensive and relevant to the target feature? What are the major differences between them, and how relevant are those differences to the target feature?

If this all sounds like hard work, it is. Arguments from analogy, when reconstructed carefully, raise many questions. If the argument has been presented in a conversational context, some of these questions can be put to the speaker, and the result can be a stimulating process of mutual clarification and discovery.

Chapter 3 Homework Exercises

Part I.
Instructions: Reconstruct each of the following arguments from analogy as a substitution-instance of the form for Argument from Analogy discussed in class. (Hint: begin by identifying the primary analogue(s), secondary analogue(s), base feature(s), and target feature.)

Example
We should expect that President Trump's administration will continue the trend of expanding the power of the central executive. After all, the Obama administration and the Bush administration before it both expanded the power of the office of the President.

Answer:
P1. The Bush administration and Obama administrations are recent presidential administrations that expanded the power of the central executive.
P2. The Trump administration is a recent presidential administration.
C. So, The Trump administration will expand the power of the central executive.

1. The professor tells us that the final exam in BIO 420 is going to be entirely short-answer. The previous two exams in BIO 420 have been entirely short-answer, and I've found them very difficult. I predict that the final exam is going to be really difficult, too.

2. Sarah and Demetrius have two young children. When asked about the experience of being parents, they say that it is exhausting but full of joy. Drew and Natalie have four children. They report similar levels of exhaustion and joy. In both of these families, both spouses work full-time, as is the case with my husband and I. Though I don't see how we

could possibly add the care of tiny humans to our already busy lives, I guess we, too, would probably find joy in the experience.

3. I was raised Catholic. I have come to believe that it's just a myth made up by humans, though. Catholicism requires that you believe in the afterlife, participate in weird religious rituals, and think there is a deity that can influence their lives. But ancient paganism involved those same things, and paganism is a ridiculous, made-up worldview.

4. Brick-and-mortar retailers should not pin their hopes on Black Friday sales this year. During the last two years, the dominance of e-commerce has meant that profits at brick-and-mortar stores have been disappointing, even on Black Friday. And if anything, we should expect e-commerce to be even more dominant this year.

5. The last two novels by Colson Whitehead, *The Underground Railroad* and *The Nickel Boys*, both explored themes of race in America. They were both brilliant and devastating. He's got a new novel coming out this year, I see, and it also explores themes of race in America. I expect it to be just as brilliant and just as devastating.

6. Today, people who publicly express "traditional" views about sexuality and gender are routinely subjected to public shaming by left-leaning voices, despite holding these "traditional" convictions earnestly and thoughtfully. This outrageous. Imagine the script were reversed: a progressive expresses her or his earnest, thoughtful conviction about, say, unjust practices by police, and is subsequently harassed and ostracized on social media. Liberals would rightly be outraged by such abuse, even though they engage in the very same reprehensible behavior.

Part II.

Instructions. Each item below is an argument from analogy, followed by a series of statements that express new information potentially relevant to the likelihood of the conclusion. For each statement, state, first, whether it (a) increases, (b) decreases, or (c) has no effect on the persuasiveness of the argument. Second, for each statement that raises or lowers the persuasiveness of the argument, state whether this is because it provides new information about (d) the quantity and relevance of the base features, (e) the number and diversity of the primary analogues, or (f) the extent of the differences between the primary analogues.

Example

Last year I planted a peach tree in my yard, and it died. Next year I will plant another peach tree in my yard. It will likely die as well.

a. Last year’s peach tree was planted in heavy clay soil, but next year’s peach tree will be planted in soft, fertile soil.
   Answer*: decreases; extent of the differences.*
b. Last year’s peach tree was planted on a Wednesday. Next year’s peach tree will be planted on a Thursday.
   Answer: *has no effect*

1. Country A recently replaced its dictatorship with a constitutional democracy. It proceeded to host free and fair elections. Country Z likewise recently replaced its dictatorship with a constitutional democracy. So, Country Z’s upcoming elections will be free and fair.

a. Majority religious, social, and artistic practices in Country A and Country Z are similar.
b. Countries B and C have likewise replaced their dictatorships with constitutional democracies, and have proceeded to host free and fair elections.
c. The population of Country A is 40 million citizens, Country B 20 million, and Country C 5 million.
d. Countries A, B, and C all have the letter ‘A’ in their name, as does Country Z.
e. Countries, A, B, and C were democracies before they were dictatorships, whereas Country Z has only ever had a monarch.
f. Countries A, B, and C have flourishing economies, as does Country Z.
g. Countries A, B, and C get a lot of press in Western media, whereas Country Z doesn’t get a lot of press.
h. The military leadership in countries A, B and C have relatively little influence, whereas the military leadership in country Z has led three coups d'état in the last 50 years.

2. Sandy’s Candies’ three stores—in Aberdeen, Bellevue, and Cashmere—have all been profitable. So, a potential new store in Darrington would also be profitable.

a. Aberdeen is a vacation town, Bellevue is a wealthy suburb, and Cashmere is a farming community.
b. The existing stores are located in shopping malls, whereas the new store would be located in a warehouse district.
c. The three stores have in common, with the proposed new location, that they are located within five blocks of an architectural firm.
d. The three stores have in common, with the proposed new location, that they are located within two blocks of an elementary school.
e. There are already two candy stores in Darrington, but there are no other candy stores in Aberdeen, Bellevue, or Cashmere.
f. The three stores doubled their profits when they started selling edible jewelry. The new store will likewise sell edible jewelry.
g. Aberdeen, Bellevue, and Cashmere all have populations whose median age is unusually low.

h. Average household income in Darrington is much lower than average household income in Aberdeen, Bellevue, and Cashmere.

Chapter 4
Inductive Generalization

4.1 Reconstructing Inductive Generalizations

Arguments from analogy are inferences from what is known about a thing or category to a conclusion about a similar thing or category. This type of reasoning is extremely common in human life because we are constantly navigating new situations, and we use our past experiences, or the experiences of those we know, to guide us.

But sometimes we shouldn't let past experiences guide us. Suppose you are considering becoming a motorcyclist. None of your three friends who are motorcyclists have ever been in an accident. So, motorcycling will be safe for you also—or so you conclude, via an argument from analogy (with three primary analogues). This is not a responsible way to make such a serious decision.

What you really need to know, in order to make an informed decision, is how frequently motorcyclists *in general* get in accidents. That is, you need statistical information about an entire category, so that you can use this statistical information as the first premise in a proportional syllogism (where the selection mentioned in the syllogism is *you.*) How would you go about acquiring this information?

One way to do so would be to go ask all living motorcyclists about the mishaps they have experienced on their bikes. But then you would spend your whole life tracking down motorcycle enthusiasts and never get around to making a decision! The good news is that there is another option: identify a *subset* of all the motorcyclists, and then generalize, on the basis of what you learn about this subset, to the whole category. Identifying the right subset is a process called 'sampling'. Let's say you're in New York City on a spring break trip. You walk up to the first 100 motorcyclists you see and ask them about their experience. 20 of them tell you that they have been in some sort of accident over the last 10 years. You conclude that roughly 20% of motorcyclists have at least one accident per decade of riding. We can reconstruct your inference as follows:

P1. Of a sample of motorcyclists consisting of 100 from New York City, 20% have at least one accident per decade of riding.
C. So, approximately 20% of motorcyclists have at least one accident per decade of riding.

This argument is an instance of an *inductive generalization*. Construed very generally, an inductive generalization is an inference from what is known about a sample of a category to a conclusion about the whole category. Inductive generalizations have the following form:

P1. Of a sample of As consisting of x1, x2, x3, etc., n% are F.
C. So, approximately n% of As are F.

Notice that inductive generalizations have just one premise! But there is a lot packed into this premise. The premise names the **population** A (in our case, motorcyclists), specifies a **sample** (x1, x2, x3, etc.) of the population, i.e. some subset of the members of that category (in our case, the 100 New York City bikers), and then states the statistical **frequency** n with which the **target feature** F (in our case, having one or more accidents per decade of riding) is found in the sample. The conclusion states that the whole population has the target feature at approximately that same frequency.

One point about the form, as it relates to the motorcycle argument. The form suggests that the sample will be specified in terms of a list of items (x1, x2, and so on). If, when you were in New York, you had taken down each biker's name, you could have done this; you could then have substituted those 100 names for the x's, resulting in a *very* long first premise:

P1. Of a sample of motorcyclists consisting of Axel, Broderick, Clovis, Drayson, Eve…[and on and on], 20% have at least one accident per decade of riding.

Usually, though, we have some shorter way of specifying which items are in our sample. And if we have a shorter way of specifying our sample, we should use it (while still including all of the relevant information).

Let's look at another example. A nurse in a clinic is instructed to determine how much plasma there is in patient X's blood. He draws a 5-gram sample of her blood. When he analyzes it, he discovers that plasma makes up 60% of the sample. He reports that the level of plasma in the patient's blood is 60%.

It is not hard to find the items that correspond to the components of the form for inductive generalization. The most obvious is the sample.

x1, x2, x3, etc.: 5 grams of patient X's blood

The sample will always be a subset of the population, so the population must be:

A: patient X's blood

The target feature is whatever feature the nurse finds at a frequency of 60% in the sample.

F: being plasma
n: 60%

We are now in a position to reconstruct the nurse's reasoning as a standard-form inductive generalization:

P1. Of a sample of patient X's blood consisting of 5-grams, 60% is plasma.
C. So, approximately 60% of patient X's blood is plasma.

(Note that "patient X's blood" shows up in our specification of *both* the sample and the population, but we don't need to include that phrase twice in the premise.)

4.2 Evaluating Inductive Generalizations

The question of whether an inductive argument is strong or weak hinges entirely on whether the sample is **representative** of the population. Two features make a sample representative. First of all, it must be **random**. In the context of sampling, randomness has a very particular meaning: *each member of the population has an equal chance of getting into the sample*. If a sample is not random in this sense, it is known as a "biased" sample. Note a similarity with arguments from analogy. Diversity among primary analogues helps make it more likely that the base features—and not some other feature that all the primary analogues happen to have in common—really do predict the presence of the target feature. Similarly, a random sample helps make it more likely that *membership in the population*—and not some other feature that all of the items in the sample happen to have in common—really does predict having the target feature (at the rate found in the sample).

Sometimes randomness is easy to establish, as it is in the case of the nurse's drawing patient X's blood. (Because we know that blood circulates through the body, drawing blood anywhere will give us a random sample.) Or again, if the population is *marbles in this bag*, simply by blindly reaching into the bag I can acquire a random sample. But sometimes randomness is much harder to establish. Think again of the 100 motorcyclists. This sample is certainly not random, since only motorcyclists in New York City had a chance at making it in. How likely is it that riding in New York is exactly as dangerous as riding elsewhere in the country? Presumably, not likely at all.

Second, for a sample to be representative, it must be **big enough**. A sufficiently large sample size is easy to come by when the population is either very small or very uniform. An example of a very uniform population is *2005 Honda Civics*. This category exhibits some diversity, for sure (with respect to color, body style, special features). But if I want to know, for example, whether 2005 Honda Civics handle well, a random sample of ten vehicles would put me in a position to make a very strong inductive generalization.

An example of a very small population is *humans who have walked on the moon*. Suppose I want to know whether humans who have walked on the moon found it exciting. The population consists of twelve individuals. A random sample of ten would, again, put me in a position to make a very strong inductive generalization.

Thus, when we are trying to learn about a population that is either small or uniform, it isn't very difficult to ensure that our sample is representative; we simply need to consult our common sense. We'll call inductive generalizations that rely on common-sense sampling 'informal inductive generalizations'. (I'll have more to say below about what I mean by this designation.)

If, however, the population is either very large or very diverse, a much larger sample is required. Suppose, for example, you want to know what percentage of U.S. citizens are Lutherans. If you were to randomly poll ten U.S. citizens, there's a good chance you would not have a single Lutheran in the mix. It's clear that you would need to poll far more. But how many more?

Here we need some help from the study of statistics. In order to understand what statistics teaches us about sampling, we will first need to learn two technical concepts used by statisticians: *margin of error* and *confidence level*. **Margin of error** is the degree of precision with which we generalize from the sample to the population. Suppose we find that 30% of the items in a sample have some feature. This doesn't mean that *exactly* 30% of the population has the feature, but rather that *approximately* 30% of the population has the feature. But what does "approximately 30%" mean? Anywhere from 20% to 40% (a range of 20 percentage points)? 29%-31% (a range of two percentage points?) The margin of error is the numerical specification of this range of approximation. It is usually stated as "plus-or-minus" some number, e.g. '+/- 2%'.

Confidence level is the probability that the population really does have the feature at the same frequency that the sample does—within the margin of error. In other words, it is a measure of the degree of inductive strength of an inference *from* some precise finding about a sample (e.g. that 30% of the items in the sample have the target feature) *to* an estimation about a population (e.g., that 30% +/- 2% of the items in the population have the target feature).

Confidence level and margin of error are related to each other in the following way: the narrower the margin of error, the less confident we can be that the population really *does* exhibit the target feature at the same rate as the sample. If you want your inductive generalization to be really strong, you have to be satisfied with a not-very-precise generalization. If you want a precise number in your conclusion, you will have to sacrifice inductive strength.

So, when inductive generalizations use statistics, there are three numerical variables in play: the size of the sample, the confidence level, and the margin of error. Statisticians have come to the consensus that what matters most is confidence level (or, in our vocabulary, that the inference be *strong*). There are mathematical formulae that researchers can use to determine, based on the size of the sample, what the margin of error will be if we insist on a 95% confidence level. We won't be learning these formulae. Instead, we'll just use a chart that captures some of this information:

Sample Size	Margin of Error (at 95% confidence level)
25	+/- 22%
50	+/- 14%
100	+/- 10%
250	+/- 6%
500	+/- 4%
1000	+/- 3%
1500	+/- 2%

Now that we understand the concepts of margin of error and confidence level and how they relate to sample size, we can return to our discussion of sampling. Recall that we were trying to determine how large a sample needs to be, when the population is very large and very diverse. We are now in a position to answer that question. First of all, we can see from our chart that the bigger the sample, the more precise of a generalization we can make. But, second, that relationship (between sample size and precision) isn't linear; it levels off, the bigger the sample size. If you start with 500 items in your sample and add another 500 items (so now your sample has 1000 items in it), you have shrunk your margin of error by 4 percentage points (from a range of 12 to a range of 8). But if you add *another* 500 items, you only shrink your margin of error by 2 percentage points (from a range of 8 to a range of 6). Returns continue to diminish as the sample size increases: adding 500 more items to a sample of a million will make virtually no difference to the margin of error.

Hidden beneath all this math is a rather surprising discovery. If you are dealing with a population that is very large and very diverse, a (random sample) of 1000 will give you pretty good information (within a range of 6 percentage points) about the entire population. To make this concrete: a political strategist can find out what percentage of Americans, within a 4% margin of error, would vote for her candidate *by asking only 1500 people*. In fact, most pollsters use a sample of between 1000 and 1500 people. (The next time you read the published results of such a poll, look at the fine print. Details about sample size and margin of error should be, and usually are, included.)

Inductive generalizations that rely on the mathematics of statistics are "formal inductive generalizations," in contrast with *informal* inductive generalizations, which do not rely on statistical methods. Recall: informal inductive generalizations are appropriate when the population is small or uniform; in those cases, we can rely on common sense to tell us how many items we need in our sample for the sample to be big enough. But when a population is very large and very diverse, a formal inductive generalization is called for. If a speaker makes an inference about a very large, very diverse population, but *doesn't* use any statistical methods in the process, then the inference is almost certainly a weak one. The simplest way to tell formal from informal generalizations is to ascertain whether a margin of error is specified. Informal generalizations can get away with the vague term 'approximately' in the conclusion. But formal generalizations *replace* this term with a precise, numerical specification of the margin of error.

In sum: inductive generalizations are strong only when the sample is representative, where this means that the sample is both (a) random and (b) big enough. In everyday reasoning, we often generalize from pretty small samples, and this is not a problem if the population is small or uniform. These inferences are *informal* inductive generalizations. But when we want to know about large and diverse populations, we need to use statistical methods. In particular, we need to generalize to the whole population only to the degree of precision warranted by the size of our sample. Such inferences are *formal* inductive generalizations. While they don't show up in everyday reasoning, they are used constantly by political and industry analysts, advertisers, psychologists and sociologists, and many other professionals.

<u>Chapter 4 Homework Exercises</u>

Part I.

Instructions: Each of the following contains an inductive generalization. (A) Specify whether it is a formal or informal inductive generalization. (B) Reconstruct it as a substitution-instance of the form discussed in class. (Hint: begin by identifying the population, sample, and target property.) (C) Specify whether the inference is strong or weak. If it is weak, state whether the weakness stems from a sample size that is too small, biased, or both. (For formal inductive generalizations, assume that inductive strength requires confidence level of 95%, and use the table in this chapter as a guide.)

Example
A consulting firm polled 100 fast-food employees in Montgomery, Alabama. When asked, "On a scale of 1 to 5, where 1 is not at all supportive and 5 is very supportive, how would you rate your employer's parental leave policies?", 75% gave an answer of 1, 2, or 3. We can conclude that between 65 and 85% of workers in Montgomery regard their employers as unsupportive of unsupportive of parental leave.

Answer:
(A) Formal

(B) P1. Of a sample of Montgomery workers consisting of 100 fast-food employees, 75% believe their employers unsupportive of parental leave.
C. So, 75% (+/- 10%) of Montgomery workers believe their employers unsupportive of parental leave.

(C) Weak: biased sample

1. I have had pizza from Johnny O's five times over the past two years, including three different pizzas they make. It was solid pizza every time. I conclude that Johnny O's pizza is pretty much always good.

2. The library staff pulled 100 books off the libraries shelves at random; 30 of them contained underlining or highlighting. So, at least 20%, and perhaps more, of all the books in the library have similar markings.

3. To test the algae content in the lake, a biologist took a ten-gallon sample of the water at one end. Algae made up 3% of the volume of the sample. Therefore, algae make up approximately 3% of the volume of the lake.

4. As they checked out of hotels, bed and breakfasts, and hostels, 1500 tourists were briefly questioned about Montreal's subway system. 79% stated that they were very satisfied with the subway system. Therefore, between 77 and 81% of all tourists are very satisfied with Montreal's subway system.

5. To estimate public support for a new municipality-funded convention center, researchers surveyed 100 homeowners in one of the city's fashionable neighborhoods. They found that 89 percent of those sampled were enthusiastic about the project. Therefore, we may conclude that somewhere between 84% and 94% of the city's residents favor the convention center.

6. The State University chapter of the American Medical Student Association wanted to know how many students at State University use the school's recreational facilities. To find out, they contacted 1,000 randomly selected State University students and asked them how often they used any of the recreational facilities on campus. Of these students, 380 said that they used the recreational facilities at least once per week. So, between 36 and 40% of SU students use the school's recreational facilities at least once per week.

7. When a random sample of 600 voters was taken on the eve of the presidential election, it was found that 51% of those sampled intended to vote for the Democrat and 49% for the Republican. Therefore, the Democrat will receive somewhere between 50% and 53% of the vote—enough to win!

8. The Hospitality Division of the Garcia-Brown Hotel wanted to determine how satisfied their customers were with the hotel's towels. It turns out that 72% (+/-4%) of customers are satisfied. Five hundred customers were randomly selected as they checked out of the hotel and asked how they had liked the towels. Of these customers, 360 said they were satisfied with the hotel's towels.

Part II. For each of the following, state whether it is a proportional syllogism, an argument from analogy, an informal inductive generalization, or a formal inductive generalization. Then reconstruct it according to its proper form.

Example
In a beta-test of Dragnet Games's latest M-rated racing game, "Slaughter Race," 6 out of the 12 players in the test-group said the game was more frustrating than fun. So, if the game were released in its present form today, about half of all players would find it pretty frustrating.

> Answer:
> *Informal Inductive Generalization*
>
> *P1. Of a sample of Slaughter Race players consisting of 12 players, 50% found the game more frustrating than fun.*
> *C. So, approximately 50% of Slaughter Race players would find the game more frustrating than fun.*

1. Ten minutes ago, someone pulled the fire alarm here at a high school. I'd be willing to bet that it was a false alarm, since 85% of the time that someone pulls a fire alarm in a high school, it's a false alarm.

2. Market research shows that 75% of the 500 Meijer customers randomly surveyed throughout the Great Lakes Region respond favorably to a friendly greeting when they enter. So, we can conclude that 75% (+/- 4%) of Meijer customers in the Great Lakes Region have a favorable response to being greeted when they enter the store.

3. Market research shows that 75% of Meijer customers in the Great Lakes Region respond favorably to a friendly greeting when they enter. This customer I see entering the Valparaiso Meijer is from that market region. Hence, she'll most likely respond favorably to my greeting.

4. In 2008, housing prices were rising, employment was above 95%, and investors were reckless. All those conditions hold today. As we all know, those conditions preceded a recession in 2008. Analogously, we should expect that a recession will soon it.

5. Over the past 30 years, there have been six major, federally-funded initiatives to improve public schools in the greater San Antonio area. Every single one of these has failed to make a measurable difference. The city council is now proposing its new Gold Ribbons for Overall Student Success (GROSS) initiative. Like its successors, it is a major, federally-funded initiative. Like its successors, it will fail miserably.

6. I have inspected three pallets' worth of the latest shipment of 6-foot cedar boards. Each pallet contained about ¼ warped boards. I conclude that about a quarter of the whole shipment of cedar boards is warped.

7. 25/31 days in December in northwest Indiana typically reach temperatures above freezing. Darla and Daryl's wedding will be held in December in northwest Indiana, so probably it will be above freezing on their wedding day.

8. In a dozen crash-tests of Honda's revamped Civic hatchback, crash-test dummies were basically unharmed, when the collisions occurred at under 40mph. Crash test dummies are designed to mimic human physiology when subjected to decelerating forces. We can conclude that human drivers of new Honda Civic hatchbacks will be likewise be safe, in collisions of up to 40mph.

Chapter 5
Informal Causal Inference

5.1 Reconstructing Informal Causal Inferences

Human beings wish to understand the world they inhabit. This desire to understand is the reason universities exist in the first place, and it is why generation after generation of philosophers, scientists, historians, and other practitioners devote their lives to the pursuit of the truth.

But we do not merely want to understand the world for its own sake: we also want to be able to alter the world to meet our needs. We want to be able to build durable dwellings, grow food, cure and prevent diseases, reduce crime and poverty, educate our children, improve relations among societies. In order to achieve these ends, we need to know what works and what doesn't. We need *causal knowledge*: knowledge of what causes what.

Let's start with a mundane example. Suppose you have just been hired as house manager of a performing arts space. It will be your job, among many other things, to turn on and off the house lights at the appropriate times during a performance. There is a bank of hundreds of light switches at the back of the stage. You flip each switch, one by one, and see what happens. Various switches seem to control various electrical circuits in the building. And then, about halfway down, you find what you're looking for: you flip switch #47, and the house lights turn off. You flip it back, and they come on again. You try it one more time just to make sure, and the same thing happens. Good, then! You're ready for the dress rehearsal tonight.

Let's think about how you made this discovery. One thing is for sure: you didn't directly *observe* the causal link between the flip-switching and the light-flickering. You just observed a couple of events and noticed that they were correlated. You noticed, in fact, that these events were correlated in two different ways. First, you noticed that they were *positively* correlated: the light-bulb turned on (or off) when the switch was flipped. Second, you noticed that they were *negatively* correlated: the light-bulb didn't turn on (or off) when the switch wasn't flipped. We can reconstruct your reasoning as an argument with two premises: one that captures this positive correlation, and one that captures this negative correlation:

P1. At 2:05pm and 2:06pm, the light turned on when switch #47 was flipped.
P2. At 2:07pm and 2:08pm, the light didn't turn on when switch #47 wasn't flipped.
C. So, flipping the switch causes the light to turn on.

This reconstructed "argument" is a rudimentary example of a *causal inference.* Construed very generally, a causal inference is an inference from observed correlations among things, categories, or events to a conclusion about a causal relationship between those things, categories, or events. Causal inferences come in many varieties. Sometimes they focus on particular phenomena (say, when a historian is trying to identify the cause of a war, or a plumber is trying to identify the cause of a leak); sometimes they are about very general phenomena (say, when an economist is trying to identify the causes of recessions, or a sociologist is trying to identify the causes of poverty). Sometimes they require little more than common sense, and sometimes they require complicated statistical reasoning. In this section we'll focus on *informal* causal inferences, which require little more than common sense. In the next chapter we'll focus on *formal* causal inferences, which require statistical reasoning. (Their forms are a bit different from one another.)

Informal causal inferences have the following form:

P1. On occasions e1, e2, e3, etc., y1, y2, y3, etc. are E when x1, x2, x3, etc. are C.
P2. [On occasions c1, c2, c3, etc., y1, y2, y3, etc. are not E when x1, x2, x3, etc. are not C.]
C. C. So, X's being C causes Y's being E.

The first premise expresses a positive correlation between two phenomena. This premise is called the "experimental premise." The second premise expresses a negative correlation between two phenomena, and it is called the "control premise." (The terms 'experimental' and 'control' are derived from formal causal reasoning, which we will cover in the next chapter. The basic idea is that when, for example, you flip the switch, you're *experimenting*—i.e. actively intervening—on the system. And when you don't flip the switch, you're "controlling for" other potential causes of the light turning on. For now, don't worry too much about what the terms mean; just try to remember the positive correlation goes in the experimental premise and the negative correlation goes in the control premise.) The conclusion then expresses a causal relationship between the two phenomena.

Both the experimental and control premises mention an **agent or agents** x1, x2, x3, etc. and a **patient or patients** y1, y2, y3, etc. Agents are the things that are allegedly doing the causing, and patients are the things that are allegedly being affected. (The words 'agent' and 'active' are etymologically related, as are the words 'patient' and 'passive'.) They also mention an **alleged cause** C and an **alleged effect** E. These are the particular features of the agent and patient in virtue of which those two items are causally related to each other. In our example above, there is only one agent—the switch—and only one patient—the bulb. The alleged cause is *being flipped* and the alleged effect is *turning on*. The switch and the bulb stand in a causal relationship to one another by virtue of particular features they can each have: being flipped (in the case of the switch) and turning on (in the case of the bulb).

The other items mentioned in the first two premises are "occasions." In the first premise, these are **experimental occasions** e1, e2, e3, etc., and in the second premises, these are **control occasions** c1, c2, c3, etc. Experimental and control occasions are simply the occasions on which a correlation—positive or negative—is observed to occur between the alleged cause and the alleged effect. In our example above, the experimental occasions are 2:05 and 2:06 (when you flipped the switch) and the control occasions are 2:07 and 2:08 (when you didn't).

The conclusion states that C, when exhibited by **the agent or agent-population** X, causes E in **the patient or patient-population** Y. Sometimes the conclusion just expresses that this relationship holds between the agent and the patient, as is the case in our example above. But sometimes it says something more general: that the relationship holds between, on the one hand, an entire population of which the agent is a member, and on the other hand, an entire population of which the patient is a member.

Let's look at another example, in which agent- and patient-populations are mentioned in the conclusion. Marvin is a dog trainer. Early in his career, he often observed that when he directed sharp, high-pitched commands to his dogs (dozens of them, of many breeds), they learned to interpret and follow the commands more quickly, but when he used slurred, lower-pitched, commands, the same dogs tended to be confused. He concludes that by giving dogs sharp, high-pitched commands, humans can speed up the training process.

We can easily identify the alleged cause and the alleged effect:

C: using sharp, high-pitched commands
E: being more quickly trained

Among what things does Marvin observe these factors to be correlated? The answer to this question gives us the agent and patient(s):

x: Marvin
y: dozens of dogs of many breeds

And on what occasions did Marvin observe these correlations? All we are told is "often," and "early in his career," so that's what we'll put both for experimental occasions and control occasions:

e1, e2, e3, etc.: often, early in Marvin's career
c1, c2, c3, etc.: often, early in Marvin's career

We can now construct our two premises:

P1. Often, early in Marvin's career, dozens of dogs of many breeds were more quickly trained when Marvin used sharp, high-pitched commands.
P2. Often, early in Marvin's career, dozens of dogs of many breeds were not more quickly trained when Marvin did not use sharp, high-pitched commands.

Now, what about the conclusion? We already know what to put in place of C (using sharp, high-pitched commands) and E (being more quickly trained). What about X and Y? The thing to notice is that Marvin doesn't just reach a conclusion about a causal relationship between *his* commands and the *particular* dogs he used those commands on. Rather, he generalizes to a causal relationship between *populations* of which he, and those particular dogs, are members:

X: Humans
Y: Dogs

The conclusion is now easy to reconstruct:

C. So, humans' using sharp, high-pitched commands causes dogs to be more quickly trained.

5.2 Evaluating Informal Causal Inferences

Before we turn our attention to what makes an informal causal inference inductively strong or weak, let's consider what makes such inferences *inductive* at all. The reason can be put in the form of a slogan: *correlation is not causation*. Take any two phenomena, A and B, that are positively correlated. There are four possible explanations of their correlation:

1. A causes B.
2. B causes A.
3. Some third thing, C, causes both.
4. Their correlation is a total coincidence.

It's easy to think of a case that could turn out to be any of these. Let's say that Prof. Goggans notices a trend in his Ancient Philosophy course: students attend the fall section regularly, and the class earns an average grade of an A-; student attend the spring semester much less consistently, and the class's average grace is lower. Here are four explanations, consistent with our four options above:

1. Students' regularly attending causes them to earn better grades.
2. Students' earning better grades causes them to regularly attend. (Maybe struggling students 'check out' and so stay home.)
3. The roster in the fall section included more students who were better motivated to learn, and this motivation caused both their better attendance and their high grades.
4. The higher grades in the fall were caused by greater enthusiasm about the material; lower attendance in the spring was caused by student illness.

For a causal inference to be strong, the premises have to make it likely that possibility #1 is the correct explanation of the observed correlations, and less likely that possibilities #2-4 are correct. How is this done?

First of all, the inference needs to include a control-premise. And very many causal inferences people make every day do *not* include this premise. That is, people make bad inferences from positive correlations to causal connections on a regular basis. Even when people arrive at correct causal principles in this fashion, they are really only making a lucky, one-out-of-four guess. (Note: if you find yourself reconstructing an informal causal inference and you can't seem to find a negative correlation mentioned anywhere, this means that you should simply omit the control-premise in your reconstruction.)

Second, the experimental and control occasions need to be *representative*. To understand what this means, let's point out one more quirk of our original example (about the switch and the bulb). The premises mention particular occasions on which correlations are observed to hold. We might say that these correlations are "time-indexed." But this time-index drops away in the conclusion. It is express an *a*temporal, ongoing causal relationship between switching-flipping and bulb-illuminating. We hadn't noticed it before, but there is an implicit inductive generalization embedded in the inference, a generization from a sample of occasions, to *any* occasion whatsoever.

Recall that for a sample to be representative, it has to be *random* and *big enough.* Our set of experimental and control occasions comprise 'samples' of a sort (drawn from the population of *all occasions*). Making these samples 'random' means making sure that the experimental and control occasions aren't unusual in some way. For example: suppose you observed, at 10:00am on weekdays, that the Valparaiso University Campanile chimes when people walk into the Chapel, and that at 10:05am on weekdays, the Campanile doesn't chime when people *don't* walk into the Chapel. You conclude that people entering the chapel causes the Campanile to chime. Something has gone wrong. What is it? The answer is that your experimental-occasions are not at all representative. There are plenty of times when the positive correlation *doesn't* hold. The beginning of morning Chapel is a special time when chimings and Chapel-enterings coincide.

Making sure that the 'sample' of occasions is *big enough* just means that the more frequently a correlation is observed to hold, the more likely it is an indication of a causal relationship. In the example of Prof. Goggans's two Ancient Philosophy sections, he would need to observe the correlation between attendance and average grade for many more semesters, or consult many other professors' records of that correlation, before he would really be in a position to infer that one causes the other.

The astute reader will ask at this point: if a correlation between phenomena A and B is observed to hold *on representative occasions*, this will help render explanations 3 and 4 really unlikely—i.e., that there is some third thing C that causes both A and B, or that the correlation is a coincidence. But how do we show that A causes B rather than that B causes A?

Let's return to our example. How do you know that flipping the switch causes the light to turn on, rather than that the light's turning on causes you to flip the switch? The answer is that you're *introducing* the causal factor, of your own free will. You know perfectly well that nothing is causing you to flip the switch.

Now, it's not always the case that we can introduce the causal factor, or even if we can, that we should. (We will return to this point in connection with scientific causal studies, in the next section.) Thus, Prof. Goggans is probably not in a position to determine whether attendance impacts grades or grades impact attendance, even if he observes the correlation over many semesters, because he isn't introducing the alleged causal factor. (He isn't forcing students to stay home.)

There is one final source of strength and weakness of informal causal inferences that we need to discuss. We noted that informal causal inferences implicitly include a generalization, from a sample of occasions to all occasions. Sometimes they include a *second* implicit generalization, when the conclusion mentions agent-population and/or patient-population, rather than the agent(s) and patient(s) mentioned in the premises. Marvin's inference was an example of an implicit generalization of this sort. Even if the correlations in the premises are robust enough to warrant an inference to *a* causal conclusion, they may not warrant an inference to a causal conclusion about populations as general as those the speaker mentions. For example, suppose we learned that all of

Marvin's dogs are medium-sized breeds. We probably aren't entitled to conclude that a chihuahua or a Great Dane would benefit from sharp, high-pitched commands. In short: if the conclusion generalizes to the agent-population and/or the patient-population, the agent(s) and patient(s) mentioned in the premises had better be representative of these broader categories. If not, the inference is weak.

We have discussed four sources of strength for informal inductive inferences. Let's briefly summarize them.

- First, the argument must include a control-premise. An inference from a positive correlation alone to a causal conclusion is weak.
- Second, the occasions must be representative. If they are not, the correlations observed are likely as not explainable in terms of a 'common cause' or a coincidence.
- Third, the argument is much stronger if the causal factor has been intentionally introduced, rather than passively observed. Otherwise, we have no way to tell whether we're getting C and E backward.
- Fourth, the agents and patients mentioned in the premises need to be representative of the agent- and patient-categories mentioned in the conclusion.

Chapter 5 Homework Exercises

Instructions: Each of the following contains an informal causal inference. Reconstruct it as a substitution-instance of the form discussed in class. Then state whether it is strong or weak. If it is weak, state whether this is because (a) it lacks a control-premise, (b) the experimental- or control-occasions are non-representative, or (c) it generalizes to an overly broad agent-category or patient-category (or some combination of these three).

Example
I have reached the conclusion that reading some of *Miracles* by C.S. Lewis helps me write more elegant prose. Here's my evidence. This past year I wrote 8 essays, all in different academic disciplines. When I wrote the first four, I found myself stumbling over my words, putting things awkwardly. For each of the second four, I began by reading two pages out of C.S. Lewis's book *Miracles*. I found that my writing was much more elegant.

> Answer:
> *P1. During the past year, four of my essays are more elegant when* Miracles *is read by me prior to writing them.*
> *P2. During the past year, four of my essays are not more elegant when* Miracles *is not read by me prior to writing them.*

So, Miracles' *being read by me prior to essay-writing causes my essays to be more elegant.*

Strong

1. Last week, sales representatives Igor, Laura, and Amelia all implemented the practice of greeting customers enthusiastically. (Before, they would say something much blander, such as, "Hi, welcome to Kohls.") All three reported a noticeable uptick in clothing sales: from three sales per day on average to five sales per day on average. So it's probable that enthusiastically greeting our customers will make a difference to our sales numbers.

2. I can remember three times when I ate a chocolate bar and my acne got worse the next day. No more chocolate for me—it causes me to have acne!

3. My eucalyptus plant seems to grow best in the shade. For the first two months after I bought it, I kept it near the window, and it didn't grow at all. When I finally put it in a dark corner of the room, it started to grow, and that's been happening for about a month now.

4. If you can't find a parking spot, just say the Lord's Prayer. It totally worked for Saul once.

5. During Obama's second term, Orville's Guns sold ten or more rifles per week. Since Trump has been in office, they have sold on average fewer than ten rifles per week. When there's a Democrat in the oval office, this causes people to buy more guns, apparently.

6. Strauss Pharmaceuticals used to award bonuses to managers who could demonstrate increased efficiency in their departments. The trouble was that, over the final year this policy was in place, employees from three different highly efficient departments reported being asked to do something unethical in the name of efficiency (such as fudge the numbers on a report or ignore safety protocols). So, upper management ended the policy. And in the year since the policy was ended, there have been no such reports.

Chapter 6
Formal Causal Inference

6.1 Reconstructing Formal Causal Inferences

In this previous section, we discussed a form of reasoning we use in everyday settings to discern the causal structure of the world, on the basis of observed correlations. We called this form of reasoning "informal causal inference."

There is a very similar form of reasoning used by scientists. In keeping with informal causal inference, *formal* causal inference starts from premises that express positive and negative correlations. Also in keeping with informal causal inference, formal causal inference implicitly includes inductive generalization. But formal causal inference makes use of rigorous statistical methods, about which we learned in Chapter 4.

An experiment that uses formal causal inference is often called a "scientific study" or a "causal study" or just a "study". (If you ever hear someone say that "studies have shown" something or other, they are probably referring to experiments of the sort we're about to discuss.) The basic set-up of a causal study is very simple. Suppose you want to test whether some causal factor C is responsible from some effect E in a population. Take a sample of that population; divide it into two groups, the "experimental group" and the "control group." Add the causal factor C to the experimental group, and then wait to see to what extent E shows up, using E's frequency in the control group as a baseline for comparison.

As an example, suppose that some researchers want to determine whether reading aloud to children helps their reading skills improve. They start by recruiting 1000 elementary-aged children from various ethnic groups and socioeconomic backgrounds whose parents do not (prior to the commencement of the study) read to them for more than a few minutes per week. These children are divided into two groups. Half of them are then read to by their parents for 30 minutes or more for three months. The other half, during that same three months, receive the same, minimal amount of aloud reading that they received before the study began. At the end of the three months, 50% of the children in the first group exhibit measurable improvements in reading, whereas only 30% of the children in the second group exhibit such improvements. The researchers conclude that reading aloud to children causes their reading skills to improve.

It will be easiest to reconstruct this argument after we have introduced the form for formal causal inference. Here it is:

P1. Of a sample of As consisting of x1, x2, x3, etc., n% are E when C.
P2. Of a sample of As consisting of y1, y2, y3, etc., m% are E when not C.
C. So, C causes E in As.

The premises state the results of the study in each of the two groups. The first premise names the **population** A (in our case, elementary-aged children); it describes the **experimental group** (x1, x2, x3, etc.), a sample of that population (in our case, 500 children from various ethnic and socioeconomic backgrounds whose parents were not much in the habit of reading to them); and it states the statistical **frequency** n with which the **alleged effect** E (in our case, measurable improvement in reading) accompanies the presence of the **alleged cause** C (in our case, being read to for 30 minutes or more each day for three months).

The second premise has the same structure, but it tells us about a different sample of the population, the **control group** (y1, y2, y3, etc.): in particular, it tells us the **frequency** m with which the alleged effect accompanies the *absence* of the alleged cause. And the conclusion, finally, states that C and E are causally related, within the relevant population. Note that C and E can sometimes pick out a more general **alleged cause-*category*** and **alleged effect-*category*.** This is because the intervention made on an experimental group has to be very specific, and we want to draw a causal lesson that applies in other contexts. Just such a generalization occurs in our reading-study: the researchers make a conclusion about reading aloud to children, not about reading aloud to children *for at least 30 minutes per day*. (As with informal causal inferences, formal causal inferences can thus contain *two* implicit generalization at once: (1) from the experimental/control groups to the whole population; (2) from specific causes/effects to more general categories of causes/effects.)

To reconstruct the reasoning used by the researchers in the reading study, we need to identify the items corresponding to teach of the variables in the form, as follows:

A: Children
C: being read to for 30 minutes or more each day for three months / being read to
E: measurable improvement in reading
x1, x2, x3, etc.: 500 elementary-aged children from various ethnic groups and socioeconomic backgrounds.
y1, y2, y3, etc.: 500 elementary-aged children from various ethnic groups and socioeconomic backgrounds.
n: 50
m: 30

(Notice that C in the conclusion takes the form of the alleged cause-*category* of 'being read to'—a more general category than the specific type of intervention mentioned in the premise.) Putting this all together, here's what we get:

P1. Of a sample of children consisting of 500 elementary-aged children from a wide variety of ethnic and socioeconomic backgrounds, 50% improve their reading skills when read-to for at least 30 minutes per day for three months.
P2. Of a sample of children consisting of 500 children from a wide variety of ethnic and socioeconomic backgrounds, 30% improve their reading skills when not read to for at least 10 minutes a day.
C. So, being read to causes reading skills to improve in children.

6.2 Evaluating Formal Causal Inferences

Assessing formal causal inferences for strength is a difficult matter, because there are quite a few potential sources of strength and weakness. We will discuss four of them.

1. Is the difference in effect-levels statistically significant? The first thing we need to know is whether the difference in effect levels between the experimental and control groups is just a statistical fluke. After all, samples of a population will differ from one another in all sorts of ways. Maybe, for example, our experimental group just happened to include more children who learn to read quickly, and our control group happened to include fewer of these, and the alleged causal factor really made no difference. How can we tell?

The answer draws on what we learned about the mathematics of statistics in Chapter 4.[2] Experimental groups and control groups are samples of the population. Let's think for a minute about what it would be like to construct inductive generalizations based on our information about each of these two samples. Looking at our example: we are told that 50% of a sample of 500 children who were read to exhibit the feature *improving their reading skills,* and we are told that 30% of a different sample of 500 children who weren't read to exhibit the same feature. What does this tell us about the two populations in question, viz., children who are read to and children who aren't read to? Recall that formal inductive generalizations involve two quantifiable X-factors: *margin of error* and *confidence level.* Let's continue to use 95% confidence level as the basic threshold for inductive strength. Here, again, is our chart that relates margin of error to sample size:

Sample Size	Margin of Error (at 95% confidence level)
25	+/- 22%
50	+/- 14%
100	+/- 10%
250	+/- 6%
500	+/- 4%
1000	+/- 3%
1500	+/- 2%

This chart tells us that, because our two groups consist of 500 samples, we can be 95% confident that (1) somewhere between 46% and 54% of children who are read to will improve their reading skills; (2) somewhere between 26% and 34% of children who are not read to will improve their reading skills. Note that there is no overlap between these two percentage ranges. This means that we can be quite confident that the difference between the effect-levels in the two groups is not a fluke. The *correlation* (between being read to and reading improvement) is real. At least, it's 95% likely that it is. The finding is, as scientists say, "statistically significant." (Bear in mind: one in twenty statistically significant findings is apt to be the result of a sampling fluke. That's what '95% confidence level' means!)

But what happens when there *is* an overlap? Suppose that in the context of some study researchers worked with experimental groups of 100 members each, and found E (the alleged effect) in 20% of the experimental group but in only 15% of the control

[2] Let us note here that the calculations we are about to describe are somewhat simpler than those that scientists use in the context of causal studies. But the basic idea is the same.

group. Because the margin of error for samples of 100 members is +/-10%, our data predicts that E will show up at a rate of 10-30% in the population that exhibits C, and a rate of 5-25% of the population that doesn't exhibit C. There is a lot of overlap in those ranges. In particular, the part of the ranges that overlap (10 to 25, or 15 percentage points) is more than half of the total range (5 to 30, or 25 percentage points). It's actually pretty likely that there is no correlation between C and E in the general population. There might even be a *reverse* correlation. The study did not find a statistically significant difference in effect-levels, in other words.

Ideally there should be no overlap between the two ranges. A lot of overlap means that the study failed to find a correlation. What about a little bit of overlap? All we can say is *maybe there's a correlation, maybe there isn't*. Let's treat the maximum amount of overlap for such in-between cases at 1/3 of the total range. We'll say of such cases that they *might* be statistically significant.

Here, then, is the method for determining whether a study has identified a statistically significant difference in effect-levels between two groups. First determine the proportion of overlap between the two ranges by dividing the magnitude of overlap by the total magnitude of the two ranges. That is, use the following formula:

> (top of overlapping range – bottom of overlapping range) ÷ (Top of total range – bottom of total range) = proportion of overlap

Then, use the following guidelines:

> If the proportion of overlap is zero, the difference is *statistically significant*.
> If the proportion of overlap greater than zero but less than 1/3, then the difference is *maybe statistically significant.*
> If the proportion of overlap is equal to or greater than 1/3, the difference is *not statistically significant.*

In sum, in order for the information in the two premises to amount to a correlation, there has to be a statistically significant difference in effect levels between the experimental and control groups.

2. Are the experimental and control groups representative of the population? Recall that representativeness is a function of randomness and sufficient size. In our discussion of statistical significance, we just assumed that our samples were random. But it is very difficult in practice to put together experimental and control groups that are randomly chosen from the population. For one thing, you can't force anyone to participate who doesn't want to. People who for whatever reason are distrustful of scientists will, because of that very fact, rarely be studied by scientists.

The problem of sample bias is particularly pronounced in the context of psychological studies. For decades, researchers in these fields have employed college students, or people living relatively close to their universities, as their experimental subjects. To be sure, people who live near large research universities exhibit a lot of diversity. But they all come from a society that is "WEIRD": *W*estern*, E*ducated, *I*ndustrialized, *R*ich, and *D*emocratic.[3] If the population of interest is the whole of humanity, samples from WEIRD populations are seriously biased. Maybe the causal relationships that show up in WEIRD populations hold in the human population, but maybe not; any inference from these populations to people in general is going to be weak.

3. Are the experimental and control groups relevantly similar? Securing a truly random sample of the population is as important as it is difficult, as we have just seen. But equally important is that experimental subjects be randomly sorted into experimental and control groups. Consider what would happen if, for example, children in our reading-study chose which group they were in. Presumably the children who *want* to be read to—because they like books!—would choose the experimental group. But in that case, the results of the study wouldn't have isolated *being read to* as the real cause of improved reading. In all likelihood, kids who like books learn to read more quickly.

So, the best way to proceed is to randomly sort experimental subjects into the two groups, thus minimizing the possibility that some other cause will disproportionately show up in one or the other group. An experiment with this sort of set-up is called a "**randomized study**"**.** But sometimes it's not possible, for practical or ethical reasons, to set things up this way. Consider, for example, a study of the effects of smoking cigarettes. We want to know whether and to what extent smoking causes cancer and other health problems. But setting up a randomized study would require that we get a bunch of people hooked on cigarettes who weren't previously. We just can't do that. This is a case where introducing the causal factor to an experimental group is outside the bounds of research ethics.

The alternative is what's called a "**prospective study**." It proceeds as follows: find a sample of the population that already exhibits the alleged cause (e.g. smoking). Treat this group as the experimental group. Then find another sample whose members don't exhibit the alleged cause (they don't smoke.) Next, make a list of all of factors you think might be causally relevant to the effect you're interested in (cancer and other health problems, in our case), and *match* members of the experimental group with members of the control group with respect to these factors. Are there a dozen single mothers in your experimental group? Make sure there are about that many single mothers in the control group, too. Are there avid runners in the experimental group? Find some avid runners to include in the control group. And so on. If there are unmatchable participants, they'll have to be left out of the study entirely. (Sample sizes in such studies tend to be smaller, for this reason.) This process is called "**matching**". It often works pretty well, as a surrogate for random sorting. But it's not a perfect system. It requires that the researcher anticipate and "control for" factors that are relevant to the effect. But there is no way to guarantee that the two samples are relevantly similar. Maybe there is some hidden, un-thought-of commonality among the experimental subjects that explains why they exhibit both the alleged cause and the alleged effect. Thus, while prospective studies can give us pretty good evidence of causal relationships, they are always somewhat suspect. All other things being equal, randomized studies make for stronger causal inferences.

[3] Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world?. *Behavioral and brain sciences*, *33*(2-3), 61-83.

Randomized and prospective studies have something in common: they play out over time.[4] ('Prospective' means *forward-looking*.) Our reading-study took place over three months. But suppose we wanted to know the effect *over a lifetime* on being read to as children. The logistics and cost of such a study might be enough to talk us out of it. But there is another problem: we might not want to wait that long to learn about the effects of reading aloud. We certainly don't want to wait that long to learn whether smoking, or other reputedly harmful behaviors, really do cause health problems.

An alternative is to do a "**retrospective study.**" ('Retrospective' means *backward-looking*.) In these studies, the researcher first identifies a sample of the population that exhibits the alleged *effect* (e.g., health problems), treating this as the 'experimental group' (though no "experiment" is performed on it). Then the researcher finds another sample that doesn't exhibit that effect, and engages in the same process of matching discussed above until she or he is confident that relevant causal factors have been "controlled for." Then the researcher queries the two groups regarding their histories—in particular, in regard to the whether and how much each group has exhibited the alleged cause in the past. For example, a researcher might identify people who have contracted lung cancer, treating them as the experimental group; *match* the individuals in this group with relevantly similar individuals who haven't contracted lung cancer, treating this second group as the 'control group'; and finally query both groups for past tobacco usage. (Note that in these studies, the question of statistical significance will be a question about differing *cause*-levels, not differing effect-levels.) Retrospective studies suffer from the same weakness as prospective studies: similarity between the two groups is established via intentional matching, rather than random sorting. And they exhibit other sources of weakness that prospective studies don't. Because they work backward from effect to cause, they don't give us a good sense for the *degree to which* the alleged cause brings about the effect. At best, they can tell us that, in some cases, the alleged cause played *some* role in bringing about the effect. Retrospective studies are thus best treated as tentative indicators of causal relationships. All other things being equal, formal causal inferences made on their basis are weaker than inferences made on the basis of randomized or prospective studies.

4. Is there the possibility of experimenter/expectation bias? Causal studies are human activities, and as with any other human activity, participants tend to root for certain outcomes. Suppose, for example, that you have a chronic illness, heretofore untreatable. You are asked to participate in a study regarding a newly developed drug. You hope that the experimental group sees a big drop in the symptoms of the disease, and that the control group doesn't seem much of a drop. If you know which group you're in, this hope might subtly influence your behavior. If you're in the experimental group, you might start eating better, sleeping longer, exercising more, etc.—even if you don't realize that your healthy choices are motivated by your desire to see the study 'succeed'. Similar hopes and fears can influence the way that researchers manage a study and interpret their

[4] There are, in fact, some prospective studies that do not play out over time. If there are a lot of data on some population from a long time ago, it can be possible to build experimental and control groups by identifying the presence of a causal factor *at that historical moment*, and then querying the two groups regarding differing effect-levels *in the present*. Economic development is often studied in this fashion.

data. After all, a 'successful' study could help secure for the researcher such prizes as journal publications, fame, influence, grant funding, and so on.

Thus, the best causal studies are "double-blind": neither the researchers nor the subjects know, until long after the data have been collected and analyzed, which group is which (experimental or control). In a drug study, for example, participants in the control group will receive a 'sugar pill' or similar phony medication that is administered in a similar way to the authentic article. And as data on the two groups are collected, researchers will label the groups "Group 1" and "Group 2", not knowing which is the experimental group and which is the control group.

Researchers find clever ways to disguise, to the participants, which group they're in. But it is not always possible to do so. Consider again our reading study: if the children being studied know that they are part of the study at all, they will know whether they are or are not being read to. It is still possible for the *researchers* to keep themselves in the dark regarding which group each set of data corresponds to. Such studies are called 'single-blind' studies. Single-blind studies are less conclusive than double-blind studies, because expectation biases may be altering the behaviors of the subjects—thus rendering the two groups different from one another in potentially relevant ways.

Let us sum up. The strongest formal causal inferences are based on *double-blind*, *randomized studies* of experimental and control groups formed via *non-biased sampling* and that exhibit a *statistically significant* difference in effect-levels. Inferences based on studies that depart from this ideal in one way or another will be comparatively weaker. Inferences based on studies that depart from it in *lots* of ways might not even reach the minimum threshold for inductive strength (raising the likelihood of the conclusion over 50%).

It is the rare causal study that establishes a causal relationship beyond all reasonable doubt. Usually, considerable qualification (about the size of the samples, about the sampling method, about the effect-levels, etc.) is needed in order to convey an accurate sense of what the study shows. Unfortunately, nobody really likes to hear that scientists have *tentatively* shown us something. When journalists report on causal studies—and this is how the public usually hears about them—most of the details are left out, or relegated to the very last paragraph. A lot of people only see the splashy headline: "Is Tomato Juice the Next Miracle Cure?" or "How to Prevent Dementia: Steer Clear of Cell Phone Towers." The only way to tell whether we should take these alleged causal relationships seriously is to look more closely (and to know what to look for).

Chapter 6 Homework Exercises

Instructions: Each of the following contains a formal causal inference. Once you have identified the population, experimental group, the control group, the alleged cause, and the alleged effect, reconstruct it as a substitution-instance of the form discussed in class.

Then specify (a) whether it is a randomized study, a prospective study, or a retrospective study; (b) specify whether it is definitely statistically significant, maybe statistically significant, or not statistically significant (and explain your answer).

Example

In order to determine whether living with a person of a different race reduces one's racial biases, researchers studied 100 college freshman at The University of Arkansas. 50 of these students were chosen at random to share a dormitory room with a person of a different race; the other fifty were assigned a roommate of the same race as themselves. Study subjects took an "implicit bias" test before meeting their new roommate and then again 3 months into the semester. 43% of those with a different-race roommate exhibited reduced racial bias, whereas 35% of those with a same-race roommate exhibited reduced racial bias.

Answer:

P1. Of a sample of people consisting of 50 randomly-selected University of Arkansas college freshmen, 43% exhibit reduced racial bias when living with a person of a different race.

P2. Of a sample of people consisting of 50 randomly-selected University of Arkansas college freshmen, 35% exhibit reduced racial bias when not living with a person of a different race.

C. So, living with a person of a different race causes reduced racial bias in people.

(a) *Randomized study*

(b) *Not statistically significant – proportion of overlap is 20/36, which is greater than 1/3*

1. In order to determine whether caffeine intake causes increased productivity in the workforce, researchers compared 100 moderate coffee-drinkers in white-collar jobs with 100 workers in similar jobs who do not drink coffee. On their performance reviews, 70% of the coffee-drinkers were rated at "very productive" or higher, whereas only 65% of the non-coffee-drinkers were rated at "very productive" or higher.

2. In order to determine whether owning a pet increases someone's attractiveness to the opposite sex, researchers studied 200 heterosexual, unmarried subjects who were not pet-owners at the beginning of the study. Half of the subjects were given a cat or dog to take care of for one year. The other half remained petless. All subjects joined the same internet dating service and actively sought dates. By the end of the year, 57% of the pet owners had been on at least ten dates, whereas only 35% of the non-pet-owners had been on at least ten dates.

3. In order to determine whether acupuncture alleviates back pain, researchers studied 500 back-pain suffers over the span of six months. The subjects were randomly sorted

into two groups of equal size. The experimental group was given a course of acupuncture treatment (about one treatment per month); the control group was not given the treatment. At the end of the six months, 15% of the subjects in the experimental group reported substantially less back pain, compared with 6% in the control group.

4. In order to determine whether raising the sales tax causes an increase in unemployment in urban areas, researchers compared data compiled from 25 cities of varying sizes that saw a rise in unemployment over the last decade, with data from 25 cities of varying sizes that did not see a rise in unemployment over the last decade. 10 of the cities in the first group had raised their sales tax during that decade, compared with 4 of the cities in the second group.

Chapter 7
Inference to the Best Explanation

7.1 Reconstructing Inferences to the Best Explanation

Suppose we learn, via causal inference, that some causal relationship holds—for example, that eating leafy green vegetables improves liver health. Knowing such information puts us in a position to make healthier choices. But notice that it leaves a big question unanswered: *why*? How is it that leafy green vegetables have an effect on liver health? Even more valuable than causal information is the *explanation* of that causal information. Explanation is the 'holy grail' of ampliative reasoning. Or, to use a different metaphor: explanatory knowledge allows us to 'get under the hood' of the world, to know not just *what* happens but *why.*

People seek explanations in all sorts of contexts, and what counts as an explanation depends on the sorts of questions asked in those contexts. Doctors and mechanics explain adverse symptoms in terms of underlying maladies and malfunctions. (We call such explanations "diagnoses"). Detectives and prosecuting attorneys explain evidence found at the scene of a crime in terms of the actions of a suspect. Physicists and cosmologists explain the regularities of the universe in terms of very basic entities and forces. Historians explain events and trends in society in terms of social and economic forces. (Explanations in all of these domains are called "theories".) Mathematicians explain quantitative relationships in terms of a few basic rules. (We call such explanations "theorems").

These are all examples of explanatory projects carried about by specialists. But explanation is very much an everyday affair as well. Amateur chefs want to understand why a dish didn't turn out as the cook book indicated it would. Students want to understand why a paper didn't earn the grade they thought it should have. And so on.

Giving an explanation for something is not, strictly speaking, a species of *reasoning*. It is, rather, a *creative* act—an act of the imagination (albeit one that is constrained by what one believes about how the world works). But figuring out whether an explanation is *true* is indeed a species of reasoning. The subject of the present chapter is how such reasoning is constructed.

Let's consider an example. After a late night of studying, you wake up just in time to make it to class. You haven't had breakfast, you're out of cereal, and you don't have time to prepare anything. What to do? You suddenly remember: you noticed last night that there's one last piece of Aunt Cindy's blackberry pie in the fridge. You crack open the refrigerator—*and it's gone, plate and all*! You spout a string of curses in the general direction of your roommate, Shannon. Here's a reconstruction of your inference:

P1. The pie is gone.
P2. That Shannon ate the pie is the best explanation of the pie's being gone.
C. So, Shannon ate the pie.

The inference you have made is known as an 'inference to the best explanation', or an 'IBE'. (Sometimes it is also called an 'abductive' inference.) Construed very generally, an inference to the best explanation is an inference from an observed phenomenon to the best explanation of that phenomenon.

The crucial premise in your case against Shannon is the second one. To say that your hypothesis is the *best* one is to say that it compares favorably with rival hypotheses. ('Hypothesis' is another name for a proposed explanation that may or may not be true.) So, let's think about some other hypotheses you are dismissing in favor of the roommate-hypothesis. Maybe your neighbor's German shepherd—a clever beast, handy with doorknobs—snuck into your apartment while you were sleeping and ate the pie. (Call this the *canine*-hypothesis). Maybe the pie suddenly de-materialized, in a one-of-a-kind freak event. (Call this the *annihilation*-hypothesis.) Maybe a thief broke into your home, felt hungry, looked in the fridge, found and ate the pie, and then left without stealing anything. (Call this the *intruder*-hypothesis.) Maybe you're in "The Matrix" (a virtual reality indistinguishable from the real world), and one of the programmers simply deleted a file marked 'lastpieceofauntcindyspie.exe'. (Call this the *Matrix*-hypothesis.) And so on. There is literally no end to the possibilities, but we'll stop there.

Some of these hypotheses will likely strike you as extremely stupid. Good. You are of sound mind. It is still worthwhile to ask *why* they seem stupid, in contrast to the hypotheses that don't seem stupid. Philosophers of science have longed mulled over this question. The answer they have come up with is that there are "virtues" that good theories exhibit. Some lists of these **theoretical virtues** are longer, some are shorter. We'll discuss four.

Let's start with the canine-hypothesis. There is one glaring problem with it: *dogs don't dispose of their dishes when they finish eating*. But the plate is gone from the fridge along with the pie. This hypothesis thus fails to explain the known facts about the case. But the roommate-hypothesis does explain those facts. Thus, the roommate-hypothesis is more *adequate* than the canine-hypothesis. We can define the theoretical virtue of adequacy as follows: a hypothesis is **adequate** to the extent that *it explains the relevant facts*.

Second, consider the annihilation-hypothesis. At least one problem with this hypothesis is that it isn't useful. There is no way to investigate whether it happened. If it really was a one-of-a-kind event, its occurrence predicts no other occurrences, so we can't draw any interesting implications form it about how the universe will behave in the future. The annihilation-hypothesis, that is, lacks the virtue of *fruitfulness*. A hypothesis is **fruitful** to the extent that *it makes new predictions or suggests new hypotheses in other domains*. Notice that the roommate-hypothesis *is* fruitful. You know what to do with it: go present Shannon with the evidence, and see what reaction you get (guilt, defensiveness, puzzlement, etc.).

Third, what about the Matrix-hypothesis? A big strike against it is that you probably believe that you live in the real world, not the virtual world of the Matrix. Moreover, you probably hold this belief *firmly* and *centrally*. Changing your mind about this belief would be *devastating* to your worldview (unlike, say, changing your mind about whether Dominos or Papa Johns makes better pizza). The roommate-hypothesis, on the other hand, better *conserves* your background beliefs. The Matrix-hypothesis lacks the virtue of "conservatism." A hypothesis is **conservative** to the extent that *it is consistent with our background beliefs.*

Fourth, consider the *intruder*-hypothesis. Presumably you *do* believe that thieves sometimes break into houses. So it would not be utterly devastating to your worldview to adopt it. (Perhaps it would require throwing out *some* of your background beliefs, such as that home-intrusions don't occur in your neighborhood, or that thieves steal more than a piece of pie once they have broken into a home.) But the hypothesis is pretty complicated, compared to the roommate-hypothesis: it involves a sequence of events, none of which entails the others, whereas the roommate-hypothesis just involves one occurrence. The intruder-hypothesis thus runs afoul of an old principle known as 'Ockham's Razor', after the medieval philosopher (William of Ockham) who articulated it: all other things being equal, a simpler hypothesis is more likely to be true.[5] To put another way, the intruder-hypothesis lacks the virtue of *simplicity*. A hypothesis is **simple** to the extent that *it commits us to believing in a small number of things, events, or relationships.*

Now that we have encountered these four theoretical virtues (adequacy, fruitfulness, conservatism, and simplicity), we can introduce the form for inferences to the best explanation:

P1. P.
P2. H_1 is the best explanation of P.
  2a. H_1 is more **adequate** than H_2, H_3, etc.
  2b. H_1 is more **fruitful** than H_2, H_3, etc.
  2c. H_1 is more **conservative** than H_2, H_3, etc.
  2d. H_1 is more **simple** than H_2, H_3, etc.
C. So, H_1.

The first premise states the **phenomenon** P that is in need of explanation. The conclusion affirms the **primary hypothesis** H_1 that, allegedly, explains P. Premise 2 states that H_1 is the best explanation of P. The form also includes four sub-premises that jointly provide support for premise 2. Each of these sub-premises states that H_1 compares favorably to one or more **rival hypotheses** H_2, H_3, etc., with respect to one of the four theoretical virtues.

Hypotheses can be long and complicated. It would be cumbersome to write them out fully whenever an 'H' appears in the form. One way to get around this problem is to state, prior to the reconstruction, what H_1, H_2, etc. are. (E.g., "H_1: *Roommate Shannon ate the last piece of pie*.") Another option is to give the various hypotheses nicknames. That's what we did above ("the roommate-hypothesis", "the Matrix hypothesis," etc.). Using the second strategy, here's our reconstruction of the disappearing-pie inference:

P1. The pie is gone.
P2. The roommate-hypothesis is the best explanation of the pie's being gone.
  2a. The roommate-hypothesis is more adequate than the canine-hypothesis.
  2b. The roommate-hypothesis is more fruitful than the annihilation-hypothesis.

[5] Ockham's original principle was "*Entia non sunt multiplicanda praeter necessitatem*": entities must not be multiplied beyond necessity. In our example, it is not *entities* that are being multiplied unnecessarily but rather *events*.

2c. The roommate-hypothesis is more conservative than the Matrix-hypothesis.
2d. The roommate-hypothesis is more simple than the intruder-hypothesis.
C. So, the roommate-hypothesis is true.

This example is a little contrived, since it mentions exactly four rival hypotheses, one for each of the four theoretical virtues. Inferences to the best explanation are often both less and more complicated than this. Less, in that they may involve fewer rival hypotheses and invoke only one or two theoretical virtues. But they can be more complicated in that different hypotheses might win out with respect to different theoretical virtues: one of the rival hypotheses might be simpler, for example, but this mark in its favor might be outweighed by its lack of conservatism.

In practice, it can be difficult to distinguish between appeals to adequacy and appeals to conservatism. Both involve a lack of harmony between the hypothesis, on the one hand, and some other bit of information. But adequacy-considerations are more specific than conservatism-considerations. In particular, they involve the success or failure of an *explanatory relationship* between the hypothesis and the relevant bit of information. That is, if a speaker is pointing out that the hypothesis succeeds or fails to *explain* something, then the speaker is invoking the virtue of adequacy. But if all she or he says is that the hypothesis is consistent or inconsistent with something, then the speaker is invoking the virtue of conservatism. Although there is usually a right answer to the question of whether a speaker is appealing to adequacy or conservatism, there is not *always* a right answer. (For example, sometimes a speaker will point to *background beliefs* that a hypothesis *explains*. Pointing this out is equally an appeal to adequacy and to conservatism.)

7.2 Evaluating Inferences to the Best Explanation

IBEs can be stronger or weaker along two dimensions: (1) *The number of theoretical virtues the primary hypothesis has to a greater degree than its rivals*; (2) *The degree of advantage with respect to each virtue that the primary hypothesis enjoys over its rivals*. In other words, if the supporting premises 2a-2d make clear that the primary hypothesis is considerably more adequate, fruitful, simple, and conservative than the rival hypotheses under consideration, the inference is strong. If only one or two virtues is mentioned, or if the degree of advantage with respect to those virtues is small, the argument is much weaker.

There are two other crucial dimensions of assessment for IBEs. These are *completeness*-concerns, as they pertain to whether any new information significantly blunts the probabilifying force of the premises.

(3) *The number and degree of theoretical virtue exhibited more fully by rival hypotheses than the primary hypothesis*. Suppose the premises of an IBE tell us that that H_1 is much, much simpler than H_2 and H_3, more adequate than H_2, and more conservative than H_3. The inference to H_1 looks to be on pretty solid footing. But suppose we also know that H_1 is less adequate than H_3, less conservative than H_2, and much, much less fruitful than either. We're no longer in a position to confidently infer that H_1 is best and

thus likely to be true. Now, considerations that favor a rival hypothesis over the primary hypothesis are not, strictly speaking, part of the inference, so they should not be included in the reconstruction. But they are no less important than the *included* considerations (that favor the primary hypothesis) for a proper assessment of the argument. You can imagine just how complicated the assessment of an IBE can become, especially when we note that a hypothesis can exhibit the selfsame virtue in different respects (for example, a hypothesis can be remarkably simple in one way—it mentions only one event, say—but complicated in another—it involves a bunch of fancy mathematics).

(4) *The extent to which plausible rival hypotheses have been included.* IBEs compare a relatively small number hypotheses with each other. What counts as "the best explanation" among this small number of hypotheses might change if another hypothesis were added to it. Now, one cannot possibly consider every imaginable hypothesis in a single IBE, and completeness doesn't require this. What it does require is that *the most plausible* hypotheses are being considered. So, fully assessing an IBE involves asking the question, "Is there a plausible hypothesis that the speaker is leaving out?" If the answer is yes, then the argument is seriously flawed, even if the primary hypothesis is the clear winner among all the hypotheses that are mentioned.

One striking feature of IBEs, in contrast to the other inductive arguments we have discussed, is that they tend to evolve over time, as new information comes in and as new hypotheses are formulated. As an illustration of this dynamic process, let's return to the Case of the Missing Pie. We noted above that the roommate-hypothesis is a fruitful one (in contrast to the annihilation-hypothesis), in that you can do something with it: you can go present Shannon with the evidence, and see what reaction you get. Let's say you do so. Shannon calmly tells you that she was visiting her aunt in Kentucky on the night in question. "But *you* were here," she says. "Is it possible you ate the pie and but forgot that you did? Perhaps while you were asleep?"

The case against Shannon has now changed in two ways. First, the roommate-hypothesis is beginning to look thoroughly inadequate. Shannon can't have eaten the pie, because she wasn't there. Second, a new hypothesis has been put on the table: the sleepwalking-hypothesis. Now, this hypothesis might not be very promising, if you have no history of sleepwalking (that you are aware of): it's alarmingly non-conservative. ("But I'm not a sleepwalker!" you respond, and march back out of Shannon's room, frustrated and bewildered.)

Whatever its vices, the sleepwalking-hypothesis is fruitful: it predicts that you'll somnambulate again at some point. And so you set up a video camera near your bed, hit record, and go to sleep. Each morning for a week you upload the video and quickly scroll through it. Sure enough: on night 7, at 2:17am, there you are getting out of bed and ambling out through the door—an activity you have no memory of. In an instant, your set of background beliefs has been updated, and the sleepwalking-hypothesis is suddenly a lot more conservative than it was previously. Shannon is not the culprit. *You are* (a conclusion you reach via an updated IBE).

The dynamic process that we just described—framing hypotheses, identifying the most plausible one, testing it and finding it wanting, considering a new hypothesis, testing it and finding that it holds up, and so on—is known as "the scientific method." Speaking very generally, it has two phases: the hypothesis-creation phase and the hypothesis-testing phase. Hypothesis-creation is an act of the imagination. Consequently, logicians have little to say about it. But the second phase—hypothesis-testing—has been very carefully scrutinized. Though empirical tests look very different in various branches of science, their logical structure is basically the same. We'll call this logical structure "the experimental method."[6]

Let's consider a simple case of hypothesis-testing in a laboratory. Suppose a geologist, Dr. Stone, is studying a mineral deposit. One of the minerals is bright red. Her hypothesis regarding the source of this pigment is that there is a substantial amount of iron oxide in the mineral. She wants to test her hypothesis. Iron oxide, she reads in her manual, absorbs a measurable amount of a certain wavelength of light. She reasons, then, that if there is a high level of iron oxide in her mineral sample (she labels it 'M'), when she puts sample M into her spectrometer, the read-out will display a non-zero number corresponding to the amount of light iron oxide absorbs.

Notice what has happened. Dr. Stone began with a phenomenon—the mineral's being red—that her hypothesis predicts. But what she wanted to do, in order to test her hypothesis, was to find some *other* phenomenon—in particular, an observable one—that her hypothesis predicts, and then go see whether this prediction holds. So she had to think about what *would* happen *were* her hypothesis correct. In other words, she has to engage in a bit of hypothetical reasoning. (We have studied hypothetical reasoning under the name 'Conditional Proof.')

She begins by assuming that her hypothesis (there's iron oxide in sample M) is true:

1. H *assumption*

But nothing of interest follows from H alone. In order to make predictions on the basis of H, she has to make some further assumptions. In particular, she assumes that *if* there is iron oxide in sample M, then sample M will absorb a measurable amount of a certain wavelength of light, and that if sample M absorbs a measurable amount of that wave-length, then a particular observable event will occur (when sample M is placed in the spectrometer, the spectrometer will display a number greater than zero.) We can conjoin these background beliefs to Dr. Stone's hypothesis, as follows:

1. $H \bullet ((H \supset S) \bullet (S \supset O))$ *assumption*

We can now represent Dr. Stone's reasoning from hypothesis to predicted observation. First, a series of simplification steps:

2. H	1, Simp
3. $(H \supset S) \bullet (S \supset O)$	1, Simp
4. $H \supset S$	2, Simp
5. $S \supset O$	3, Simp

[6] Sometimes the phrase 'scientific method' is used to refer exclusively to the process of hypothesis-testing. This is a misnomer. Running experiments is only half of the scientific endeavor. The other half is the creative task of generating testable hypotheses.

Then, a couple of uses of Modus Ponens:

6. S	2, 4, MP
7. O	5, 6, MP

Dr. Stone then discharges her assumptions by reaching the following conclusion:

8. [H • ((H ⊃ S) • (S ⊃ O))] ⊃ O

Another way to put all of this is to say that H, in connection with certain background beliefs, generates the prediction that O. Note the form that the **prediction** O took in Dr. Stone's case: 'when sample M is placed under the spectrometer, the spectrometer will display a non-zero number.' That is, it is of the form *when conditions such-and-such hold, phenomenon thus-and-so will occur*. To be testable, predictions have to be both directly detectable (usually, by one's sense organs) and utterly precise.

The background beliefs scientists use to generate predictions are called '**auxiliary hypotheses**'. Dr. Stone invoked two such hypotheses. One of her auxiliary hypotheses involved established scientific theory (that iron oxide absorbs light in a certain way). The other auxiliary hypothesis appealed to the proper functioning of her laboratory equipment (that its display will express an accurate measurement of the sample). Practically every application of the experimental method will involve auxiliary hypotheses about precisely those two things (that certain scientific laws or theories are true; that lab equipment is functioning as intended), and more: other auxiliary hypotheses might appeal to mathematical rules, commonsense beliefs, the laws of logic, the appropriateness of certain rules-of-thumb (e.g. that friction can be ignored), etc. We can represent the set of auxiliary hypotheses used on any particular occasion with an 'A' in brackets (which denote a 'set' of something). Here, then, is a schematic representation of the reasoning use to set up an empirical test:

First, assume H • {A}.
Second, from H • {A}, derive some O (by applying logical rules).
Third, conclude that [(H • {A}) ⊃ O] (by Conditional Proof).

Once an empirical test has been carefully set up in this fashion, the scientist is ready to perform it, and see what happens. There are two possibilities: the scientist observes what was predicted (this is called the "**verification**" condition) or the scientist does not observe what was predicted (this is called the "**falsification**" condition).

Let's consider the second condition first. Suppose Dr. Stone's prediction is falsified: the display on her spectrometer reads "0".

We can now add '~O' as a new premise in her reasoning:

9. ~O

The question is: what follows from this new premise? The answer is *not* that ~H follows from this premise. Rather, what follows (by modus tollens) is the negation of the hypothesis *conjoined with the auxiliary hypotheses*, viz.:

10. $\sim[H \bullet ((H \supset S) \bullet (S \supset O))]$. 8,9, MT

In other words, Dr. Stone has proven that either

- It's not the case that sample M contains substantial amounts of iron oxide, *OR*
- It's not the case that, if sample M contains substantial amounts of iron oxide, then sample M will a measurable amount of a certain light-wavelength, *OR*
- It's not the case that, if sample M absorbs any amount of a certain light-wavelength, then the spectrometer will display a non-zero number.

It is therefore consistent with the falsification of Dr. Stone's prediction *that her hypothesis was correct after all.* Maybe all that has gone wrong is that her manual misreports the particular wavelength of light absorbed by iron oxide, or that her spectrometer's display is broken, or that the instrument is not sensitive enough to detect the amount of iron oxide she analyzed. Maybe spectrometers just don't work at all; they're based on bogus science. Maybe there's no such thing as 'absorbing light'. Now, none of these possibilities might be *at all likely*. Perhaps it is overwhelmingly more probable that Dr. Stone's guess about the chemical makeup of sample M was incorrect than that any of these possibilities are actual. For our purposes, the important thing to notice is that the experiment itself doesn't tell us about these probabilities. In cases of falsification, it is up to the scientist to decide whether to toss out her hypothesis, or whether to go hunting for some other flaw in the set-up of the experiment.

Suppose, alternatively, that Dr. Stone *does* observe what she predicted. Let's exchange '~O' in line 9 for 'O':

9. O

What follows now? The answer is: *nothing*, deductively speaking. If Dr. Stone were to infer H (by Simplification from $[H \bullet ((H \supset S) \bullet (S \supset O))]$), she would be committing the fallacy of Affirming the Consequent. There is, after all, more than one possible explanation of the spectrometer's displaying a non-zero number. Perhaps it is stuck displaying that number *all the time*. Perhaps it displays numbers at random. Perhaps there is colorless, odorless gas in the spectrometer alongside sample M, and the spectrometer is actually (accurately) measuring *this* substance. Again, it could be that all of these possibilities are far less likely than that H is true. But that judgement comes from somewhere else, besides the experiment itself.

When a prediction based on a hypothesis is verified, this doesn't prove that the hypothesis is true. And when a prediction based on a hypothesis is falsified, this doesn't prove that the hypothesis is false. But this doesn't mean that empirical tests tell us nothing. On the contrary, they tell us something extremely important *about the adequacy of the hypothesis*. Consider: when Dr. Stone first concocted her hypothesis, she did so in light of a particular phenomenon, viz., the color of the mineral. With respect to this phenomenon, her hypothesis was, of course, adequate. But, so far as she knew at that early stage of investigation, it was adequate with respect to that phenomenon *alone*. But suppose she runs her test and her prediction is verified. Now her hypothesis explains *two* phenomena: (1) the color of the mineral, and (2) the reading of the spectrometer. This raises the degree to which her hypothesis exhibits the theoretical virtue of adequacy. Suppose she performs a second empirical test of a different sort, and her prediction is verified. There are now *three* phenomena with respect to which her hypothesis is adequate. The hypothesis is looking harder and harder to beat, at least when it comes to adequacy.

Of course, if the hypothesis repeatedly *fails* empirical tests—in the sense that predictions made on its basis are falsified—then its credentials with respect to adequacy will look correspondingly weak—especially if we can think of alternative explanations that are relatively fruitful, conservative, and simple.

To sum up this rather long section: inferences to the best explanation tend not to stand still. They require continual updating, as new information comes in about the virtues and vices of competing hypotheses. The experimental method is a way to accelerate this process of updating. Fruitful hypotheses do more than explain the original phenomenon that inspired them. They also make testable predictions. The experimental method is the process of generating such predictions and then seeing whether they pan out. Each empirical test is an opportunity for a hypothesis to demonstrate its adequacy (or inadequacy). Thus, 'assessing' an IBE for strength and completeness might mean more than taking stock of the static levels of the four theoretical virtues that competing hypotheses exhibit. It might mean donning a lab coat and taking up the experimental method for oneself.

7.3 Coda: On "Science"

'The scientific method' is often identified with the *experimental* method—the process of constructing and performing empirical tests. As we have noted, this is an unhelpfully narrow understanding of science, as it excludes the *creative* process of generating hypotheses in the first place. But even if we include both hypothesis-creation and hypothesis-testing in our conception of the scientific method, there is more to science than 'the scientific method'. This is because the scientific method is only relevant to two of the four theoretical virtues: adequacy (empirical tests tell us the extent to which a hypothesis is adequate, not—directly—whether it is true) and fruitfulness (one way for a hypothesis to be fruitful is for it to generate testable predictions). Conservatism and simplicity are not related to how well a hypothesis performs on empirical tests; yet these two theoretical virtues play a central role in how scientists themselves evaluate theories: they tell scientists *where to begin*. Without treating simple, conservative theories as more plausible than complicated, non-conservative theories, scientists would not be able to pare down the list of possible explanations to those that are worth studying.

Doing science means more than doing experiments, then. At its most general, it means making inferences to the best explanation, using all of the tools at one's disposal to determine which explanation is best. But this expanded picture of the scientific enterprise is at odds with popular opinion, in two ways.

First of all, there is no such thing as "scientific proof," if by 'proof' we mean *deductive* proof. IBE is an inductive argument form. It is always logically possible—if monumentally improbable—that our preferred scientific theories are false, and no amount of experimentation in a laboratory can conclusively demonstrate otherwise.

Second, there is no clear line dividing the scientific enterprise and the everyday enterprise of trying to understand the world. Yes, the experimental method is an especially rigorous way of assessing the adequacy of a hypothesis. But scientists also evaluate hypotheses in terms of whether they are consistent with our background beliefs (and not just 'scientific' background beliefs, whatever that qualifier might mean). There are empirically adequate yet crazy theories, and scientists are often right to reject them on grounds that have nothing to do with laboratory tests.

What is "science," then? There are probably a number of overlapping but distinct notions that people have in mind when they use the term: the experimental method; a set of familiar questions about the world; a more or less stable set of answers to these questions; the communities of researchers entrusted with finding these answers. The term 'science' comes from a Latin word, *Scientia*, whose meaning was very broad—something like *systematic knowledge of a subject-matter.* Modern thinkers have often wanted to *contrast* science with other ways of describing the world—morality, philosophy, religion, story-telling. While there may be lots of differences between these arenas and, say, physics, chemistry, and biology, there will be lots of points of contact as well.

Science is fallible, motley, and messy, and the border between it and other human activities is porous. Pointing this out does not discredit the extraordinary achievements of modern science in helping us understanding the world. Rather, it shows those achievements to be all the more marvelous.

Chapter 7 Homework Exercises

Part I.

Instructions: The exercises below involve a phenomenon in need of an explanation, followed by three rival explanations. Each explanation is then supplementing with additional information relevant to whether it exhibits a theoretical virtue (adequacy, conservatism, fruitfulness, and simplicity). For each bit of supplemental information, state whether it reveals the hypothesis to be better or worse, and owing to the presence/absence of which virtue.

Example

Phenomenon. Ants can easily find their way back their nests after long and meandering journeys.

Explanation: Ants have a neural "pedometer," i.e. a mechanism that counts the number of steps they take on their outbound journey. This neural mechanism then guides the ant in taking those steps in reverse on its return journey.

a. This hypothesis fails to explain how ants know which turns to make on their return journey.

b. Our best theory of the way most animals navigate involves tracking the sun's position, not the use of neural pedometers.

c. We could test this hypothesis by lengthening the legs of ants on their outbound journey and then returning their legs to normal length on their return journey. (The same number of steps would not in fact succeed in returning ants to their nest.)

d. Neural pedometers are very special neural mechanisms; believing in them would complicate our picture of ant brains considerably.

Answer:
a. worse - ADEQUACY
b. worse - CONSERVATISM
c. better - FRUITFULNESS
d. worse - SIMPLICITY

Phenomenon A. Average Grade Point Averages at Haddon Heights University have risen in the last three decades. In 1981: 2.4. In 1991: 2.7. In 2001: 2.9. In 2011: 3.2.

Explanation 1: An underground student organization has been increasingly successful in helping students cheat on tests.

a. Studies have repeatedly shown that Haddon Heights students have high standards of honor.
b. We do know that such organizations exist at many other institutions.
c. This hypothesis is more complicated than hypotheses that don't invoke some sort of conspiracy.

Explanation 2: Faculty grading standards at Haddon Heights have dropped.

a. There is no way to test this hypothesis.
b. This is inconsistent with the faculty's assertion that grading standards have remained unchanged.
c. This is consistent with the consensus among education experts, according to which most instances of grade inflation are to be explained in terms of softening grading standards.

Explanation 3: Student quality at Haddon Heights has improved.

a. High school GPA and SAT scores for entering freshmen have remained constant since 1981.
b. We can test this hypothesis by digging up some exams from the 1980s and having current students take them.
c. This is inconsistent with the consensus among education experts, according to which most instances of grade inflation are to be explained in terms of softening grading standards.

Phenomenon B. The Astoria Heights building in downtown Stewartsville caught fire at 12:29am. The fire began in a 7^{th} floor apartment and had spread to floors 7-10 by 2am, when the fire was finally extinguished.

Explanation 1: A janitor left his radio in the hallway. Lucy, a 13-year-old resident of the 7^{th} floor, dropped the radio into the janitor's bucket of soapy water, releasing an electrical current that separated the water into hydrogen and oxygen gases. The hydrogen gas was ignited by a candle in the adjacent apartment.

a. This hypothesis explains why there were traces of hydrogen gas found on the 7^{th} floor and why the janitor's radio was found in his water.
b. On the one hand, the proposed sequence of events—beginning with the dropping of the radio and the ending with the fire—is implausibly long and complicated.
c. On the other hand, the hypothesis explains a number of seemingly unrelated facts in terms of a single catalyst.

Explanation 2: An arsonist ring has been hitting the tallest buildings in each neighborhood of Stewartsville. This arsonist ring is responsible.

a. We can use this hypothesis to try to prevent future fires: we can encourage the arson division of the Haddon Heights police department to focus its attention on the tallest buildings in the city's neighborhoods.
b. Not all of the fires in the last two months have been in the tallest building in the neighborhood.
c. This hypothesis nicely explains most of the arsons over the last few months in terms of a single perpetrator.

Explanation 3: A Satanic cult that meets on the 7^{th} floor held a ritual in which fire-demons were conjured.

a. Those of us who don't already believe in fire-demons would have to revise a lot of our beliefs about the way the world works in order to endorse this explanation.
b. This hypothesis doesn't explain why there were traces of hydrogen gas found on the 7^{th} floor.
c. This hypothesis explains the odd, geometrical pattern of burn marks found in the hallway on the 7^{th} floor.

Part II.
Instructions. Each of the following contains an inference to the best explanation. Reconstruct it as a substitution-instance of the form discussed in class. (Hint: begin by identifying the phenomenon, hypothesis, and any rival hypotheses that are mentioned.) Make sure to include some or all of premises 2a-2d, depending on which theoretical virtues are alluded to.

In addition: if the inference involves an empirical test, state (a) the predicted observation, (b) any auxiliary hypotheses that are mentioned, and (c) whether the prediction is verified or falsified.

Example
Last week, I was having difficult recalling the material we are studying in my International Marketing class. I thought: maybe I'm just not any good at marketing. But that didn't seem right, since I had previously done really well in marketing classes. A second possibility that occurred to me was that I wasn't getting enough sleep. If so, I realized that there would be an easy way to tell: I could go to bed an hour earlier, and then see if I do better on the quizzes (assuming that the quizzes didn't all of a sudden get a lot easier). So, all this week, I have gone to bed an hour earlier. Sure enough: my quiz grades have been much improved.

> Answer:
> *P1. I am having difficulty recalling material in International Marketing class.*
> *P2. The sleep-deprivation hypothesis is the best explanation of my having difficult recalling material in International Marketing class.*
>> *2a. The sleep-deprivation hypothesis is more conservative than the bad-at-marketing hypothesis.*
>> *2b. The sleep-deprivation hypothesis is more fruitful than the bad-at-marketing hypothesis.*
>> *2c. The sleep-deprivation hypothesis is more adequate than the bad-at-marketing hypothesis.*
>
> *C. So, the sleep-hypothesis is true.*
>
> *Predicted observation: When I go to bed an hour earlier for a week, I earn higher quiz grades*
> *Auxiliary hypothesis: The quizzes are not easier this week than last week*
> *Observation verified*

1. A top-secret list of aliases of American spies is for sale on the dark web. That list was encrypted using the CIA's most sophisticated algorithm. Maybe a hacker has found a way to hack the CIA's servers and crack their most sophisticated codes. Or maybe someone on the inside has decided to sell government secrets. I honestly don't know which would be more frightening! Now, if this latter hypothesis were correct, we would expect that, if we check the log of CIA internet traffic, we will find a record of a CIA computer's

accessing the dark web sometime in the last month (that is, assuming that the hacker was using a CIA computer while trying to make the sale). And we *do* find that one Agent Coleman accessed the dark web seven times last month from a CIA computer. Given how complicated the CIA's cryptographic algorithms are, it would probably take a dozen networked supercomputers to crack them. So, ladies and gentlemen, I conclude that there is a mole in the Central Intelligence Agency.

2. Copyright infringement requires the intent to duplicate. Did Dave Gremlin, songwriter for The Seven Seals, intend to duplicate the chord progression used in Monster Marsh's hit single, "The Corpuscular Jugular?" Or, is the similarity between that song and Gremlin's "Rockbehind" a coincidence? Dave Gremlin, as we who have followed his career all know, is a musician of astonishing creativity. He has never *needed* to steal an idea from anyone, since he has never lacked original ideas. Furthermore, his bandmates attest that his integrity is impeccable: he would never steal musical ideas even if he had a reason. Ladies and gentlemen of the jury, since the accusations brought against the defendant are simply inconsistent with what we know about his skills and his character, we can only conclude that any similarity between the two songs in question was the result of two great musical minds thinking alike.

3. Where's my wallet? Maybe it got stolen while I was at the bar last night. It was in my purse, and I had my purse on me the whole time. Hmm. A master pick-pocket could have been watching me the whole time, waiting for a moment when I looked away from my purse, reached in, and grabbed it. He probably would have needed an accomplice to keep an eye on everybody else in the bar…No, this is getting too complicated. I must have just left it somewhere in my apartment when I got home.

4. Revelation, the last book of the Christian Bible, is a very strange book, full of images of many-headed beasts and pale horses and bowls of wrath and so on. What is going on in this book? There have been many interpretations over the centuries, but they can be roughly divided into two: the eschatological interpretation and the apocalyptic interpretation. According to the eschatological interpretation, the book consists of a set of prophecies about the events that will take place at the end of the world. ('Eschatology means "account of the last things".) According to the apocalyptic interpretation, the book is rather a re-describing (or an "unveiling", which is what the word 'apocalypse' means) of events taking place at the time of its writing in the first century A.D. In favor of the first interpretation are apparently future-directed passages such as Chapter 1 verse 19, and the famous description of the "new heaven and new earth" in chapters 21-22. Advocates of the eschatological interpretation also point to the many echoes of the Old Testament book of Daniel, the book of Daniel is widely believed to be eschatological in aim. But the trouble with the eschatological interpretation is that it simply cannot explain many of the details in the middle chapters of the book. For example, the author of Revelation says that the "number of the beast" is the number of a person—that is, the numerical sum of the numbers associated with the Hebrew letters in a person's name.

And the number mentioned—666—is the numerical sum of the letters associated with the name ‘Nero Caesar’. Thus, the apocalyptic interpretation is to be preferred.

5. Many people from all over the world who have almost died claim to have had encounters with a spiritual world (so-called “near-death experiences”). Their reports have striking similarities: a feeling of being outside of one’s body, seeing a light, encountering other disembodied persons, etc. If we want to understand these experiences as mere hallucinations and not real glimpses of the afterlife, we have to come up with some complicated story about why the reports are so similar. But, of course, no well-educated person can take the idea of the “soul” seriously anymore. Despite its simplicity, the hypothesis that near-death experiences are veridical must be rejected in favor of the hallucination hypothesis.

6. For a second there, I thought I was pregnant. See, this last week I’ve felt nauseated. I haven’t wanted to eat pizza, and I *always* want to eat pizza. And I’ve had these crazy mood swings—laughing one second, crying the next. It occurred to me that I could be pregnant. But then I bought a home pregnancy test, and the result was negative—just the one pink indicator line, not two! Assuming the pregnancy test didn’t give me a false negative, I must just be going through another bout of depression.

Chapter 8
Moral Reasoning

8.1 Introducing Moral Arguments

We human beings are not entirely at the mercy of our instincts. Rather, we sometimes make choices that are contrary to our instincts, because we know that our instincts are not always tuned to the key of morality. For example, if you are struggling with a writing assignment and have limited time, your instinct to avoid struggle and anxiety might incline you to plagiarize. But you know better, so you resist your instincts and soldier on. If at a party, an attractive but very drunk partygoer propositions you for sex, your sexual appetite might incline you to accept the proposition. But again, you know that it would be wrong to do so, so you choose contrary to appetite. These two simple cases show us that morality plays an important, guiding role in human life.

These are cases where the morally right thing to do is obvious. And that's true most of the time. But there are difficult cases, where we're either not sure what the right thing to do is, or we're pretty sure but someone we trust disagrees with us. For example, suppose you cannot afford enough food for your family, and your stingy neighbors have plenty. Should you steal from them? Suppose you have signed a non-disclosure agreement with your employer, and later discover that your company is engaged in ethically and legally dubious practices. Should you share what you know? Or again, suppose you are a low-ranking military officer, and your superior tells you that many people will die if you don't extract a crucial piece of information from a prisoner. Should you torture him? Perhaps you're not sure what to say about these cases. Even if you're pretty sure what to say, chances are that other members of your society disagree with you.

In such cases of moral uncertainty or disagreement, moral arguments are called for. Moral arguments, like all persuasive arguments, begin with premises that are better agreed-upon than the conclusion that they wish to establish. They leverage moral knowledge and consensus in order to resolve moral uncertainty and disagreement.

Moral arguments are arguments whose conclusion is a moral statement. A **moral statement** is a statement that ascribes a moral property to something, where a **moral property** is a feature that a thing, person, or action can have whereby it is valuable or disvaluable.[7] Examples of moral properties include:

---

[7] Philosophers sub-divide the category of moral properties in a number of ways. One common division is between "evaluative" (or "axiological") properties, on the one hand, and "normative" or ("deontic") properties on the other. Evaluative properties are *goodness (*or *badness)* properties, exhibited primarily by things and persons; normative properties are *oughtness* (or '*ought-not-ness*') properties, exhibited primarily by actions.

Some philosophers use the term 'prescriptive' in place of 'normative'. Others reserve the term 'prescriptive' to refer to *positive* normative properties (ought-to-be-done-ness), and contrast that category with "proscriptive" properties (ought-*not*-to-be-done-ness).

Alas, some philosophers use the terms 'evaluative' and 'normative' as basically synonymous with 'moral'. Others use these terms even more broadly, to denote value of any sort, moral or not—such as aesthetic value (e.g. beauty) or prudential value (e.g. conduciveness to

Goodness and badness
Rightness and wrongness
Being justified / unjustified
Being virtuous / being vicious
Being just / being unjust
Being an action that we ought to perform
Being a piece of legislation that we should pass

Moral properties can thus be quite simple and general (e.g. *goodness*) or extremely complicated and specific (e.g., *being the noblest way to respond to unfair criticism*). Normative statements *ascribe* such properties to something or other. Sometimes moral properties wear their evaluative component 'on their sleeve', as it were, but sometimes this component is disguised. For example, *being courageous* is a moral property, because courage is not a value-neutral psychological trait (as *being outgoing* might be, say). And there are some properties whose status as moral or non-moral is a matter of dispute: *being healthy*, say.

Moral statements *ascribe* a moral property to something. The following are examples of moral statements:

(1) Murder is wrong.
(2) Standish makes foolish financial decisions.
(3) Cruelty to animals should be punishable by law.
(4) You shouldn't listen to anything she says.

In the first statement, *wrongness* is ascribed to murder; in the second, *making foolish financial decisions* is ascribed to Standish. The third and fourth are a little trickier to parse. In the third, a moral property is being ascribed to *cruelty to animals*, but what property? The answer is *being something that should be punishable by law*. The fourth statement is trickier still, since the moral property—*being something you shouldn't listen to*—isn't being ascribed to the grammatical subject of the sentence, but rather to the grammatical object, viz., *anything she says.* Moral statements can therefore come in all shapes and sizes, provided they involve the ascription of a moral property to something. Any statement that's not moral is a **non-moral statement.**

An aside: it has become popular to contrast moral statements with *descriptive* statements. The idea is that statements such as 'John is six feet tall' *describe* the world, whereas 'John is a bad person' *evaluate* the world. The trouble with this way of talking is that the statement 'John is a bad person' seems also to *describe* John. In particular, it describes—accurately or inaccurately—his moral character as poor. Even less helpful is the distinction, popular in grade school textbooks, between 'facts' and 'values' (or sometimes between facts and 'opinions'), where evaluative statements fall into the 'value' (or 'opinion') category, and statements such as 'John is six feet tall' fall into the 'fact' category. To see why this distinction doesn't make sense, suppose that John

happiness). If an author is using any of this jargon, don't be intimidated. Just try to figure out which region of value-space the author is trying to talk about.

torments animals and vandalizes property for fun, routinely betrays his friends, manipulates people into doing what he wants, steals whenever he can get away with it, and so on. It would appear to be a fact about John that he is a bad person.[8] In short, when a statement correctly ascribes a moral property to something, that statement *describes* the world accurately and expresses a *fact*. So alleged distinctions between evaluations and descriptions, or between facts and values, are spurious.

We said above that moral arguments are arguments whose conclusion is a moral statement. There is an old saying in moral philosophy known as the 'is-ought principle' (or 'Hume's dictum', after the 18th-century philosopher, David Hume): "You can't get an ought from an is." Put in our terminology, the is-ought principle says that you can't derive a moral statement from a set of exclusively non-moral premises. This means that every moral argument will include at least one moral premise. (It is a matter of debate among philosophers as to whether the is-ought principle has exceptions. It is unlikely that you will be seeing any of the supposed counter-examples to the principle in this class. The principle is at the very least a trustworthy rule of thumb.)

To say that a moral argument is an argument with a moral conclusion (and thus, as we have just said, at least one evaluative premise) is to say that moral arguments are distinguished by their *content*, not their *form*. There is no special 'moral argument form,' in other words. Moral arguments can come in any form, deductive or inductive. Here, for example, is a moral Modus Ponens:

> P1. If sexual intercourse is a spiritual act, then casual sex is wrong.
> P2. Sexual intercourse is a spiritual act.
> C. So, casual sex is wrong.

But there are some forms that moral arguments take more frequently than others. We will be focusing on three of these forms, which we will call "moral syllogism," "moral analogy," and "moral generalization." These forms very closely map onto the first three inductive argument forms that we covered: proportional syllogism, argument from analogy, and inductive generalization. Note, though, that moral syllogisms are *deductive* arguments. In particular, they are categorical syllogisms. Moral analogies and moral generalizations are *inductive* arguments.

8.2 Moral Syllogism

Moral deliberation is often a matter of figuring out which principle applies to the situation at hand. For example, suppose that much of your backyard is overgrown with milkweed. You want to plant a garden, so you are considering uprooting the milkweed. But then your neighbor tells you this:

> Contributing to the disappearance of monarch butterfly habitats is something you shouldn't do. But uprooting the milkweed in your yard would contribute to

[8] Note well: describing John as morally bad is not the same as describing him as *worthless*. Perhaps all people are of equal worth—maybe infinite worth—whether they are morally good or bad.

disappearance of monarch butterfly habitats. So, uprooting the milkweed in your yard is something you shouldn't do.

Your neighbor has just given you a **moral syllogism.** Construed very generally, a moral syllogism is an inference from a general moral principle to a particular case that falls under that moral principle. In our example, the general moral principle is a principle about the wrongness of destroying butterfly habitats, and the particular case is the uprooting of the milkweed. You hadn't realized that the action you were considering—uprooting the milkweed—is a species of a more general *type* of action that is morally problematic, viz., contributing to the disappearance of monarch butterfly habitats.

Moral syllogisms are valid, deductive arguments. They can be paraphrased as standard-form categorical syllogisms, and then proved using a Venn Diagram. Despite the fact that moral syllogisms are deductive, they share an important conceptual similarity with proportional syllogisms (which are inductive). Proportional syllogisms, recall, make an inference from what is known about an entire category to a conclusion about a member of that category. For example,

P1. 90% of toads are brown.
P2. Mr. Toad is a toad.
C. So, Mr. Toad is brown.

The first premise is a statement about an entire category, and the second premise states that some item falls under that category. (Remember: if we replaced '90%' with '100%', the argument would be a deductively valid categorical syllogism.) Moral syllogisms are very similar: their first premise states that all of the items in a category have some moral property, and the second premise states that something or other falls under that category. Here is the form for Moral Syllogism:

P1. All As are M.
P2. x is an A.
C. So, x is M.

The first premise says that all members of some **population** A (in our example, *acts of contributing to the disappearance of monarch butterfly habitats*) have some **moral property** M (in our example, *being wrong* or *being something you shouldn't do*). The second premise says that some **selection** x is a member of the population in question. (In our example, the selection is *your act of uprooting the milkweed*—an act which, according to your neighbor, is a member of the broader class of *acts of contributing to the disappearance of monarch butterfly habitats.*) The conclusion states that that the selection has the relevant moral property (*your act of uprooting the milkweed is wrong*). Here, then, is a reconstruction of your neighbor's argument, according to the form:

P1. All acts of contributing to the disappearance of monarch butterfly habitats are wrong.
P2. Your act of uprooting the milkweed is an act that contributes to the disappearance of the monarch butterfly population.

C. So, your act of uprooting the milkweed is wrong.

There are two important things to notice about moral syllogisms. The first is that their conclusions are usually—though not always—statements about what's *prima facie* right or wrong. (Most philosophers pronounce 'prima facie' as "*pry-muh-fay-shee*".) 'Prima facie' is a Latin phrase that means "on first appearances" or "on the face of it." To say that some act is prima facie wrong is to say that there are moral considerations that make it wrong, all other things being equal. But sometimes all other things aren't equal. There can be *other* moral considerations that count in favor of it. For example, suppose it turns out that the milkweed in your yard is infested with venomous spiders. We can construct a separate moral syllogism that reaches an apparently opposite conclusion:

P1. All acts of removing infestations of venomous spiders near to where people live are morally justified.
P2. Your act of uprooting the milkweed is an act of removing an infestation of venomous spiders near to where people live.
C. So, your act of uprooting the milkweed is morally justified.

Your act of uprooting the milkweed, it turns out, is a member of multiple action-categories, and actions of these different categories might exhibit *different* value-properties, prima facie. Uprooting the milkweed might be morally justified, *all things considered* (or "ultima facie"—its *ultimate appearance*, morally speaking), even if it was prima facie morally wrong in light of your neighbor's considerations. Most speakers do not distinguish between prima facie ascriptions of moral properties and all-things-considered ascriptions of moral properties, as your neighbor did not. (One reason they do not is that it is less rhetorically powerful to qualify one's conclusion as merely prima facie.) Sometimes speakers stress that they mean the first premise to admit of no exceptions or overriding considerations. For example, "Lying is *always* wrong, no matter what" is a way of indicating that the speaker intends an all-things-considered ascription of wrongness. But absent such indicators, it is usually best to assume that the speaker is only intends a prima facie ascription of a moral property.

The second thing to notice is that moral syllogisms honor the 'is-ought principle'. They include exactly one moral premise—the first premise, i.e., the general moral principle—and exactly one non-moral premise—the second premise. In different contexts of moral disagreement, it can turn out that the real *source* of disagreement is either one of these premises. For example, suppose someone argues as follows:

> The institution of marriage should be abolished. This is because all institutions that treat women as property instead of as persons should be abolished [moral premise], and marriage treats women as property instead of as persons [non-moral premise.]

If you are unsympathetic with the conclusion—that the institution of marriage should be abolished—you might be inclined to think that the speaker has moral views that are massively divergent from your own. But once you reconstruct the argument, you might find that you actually *agree* with the moral principle from which the speaker is arguing.

We should indeed abolish institutions that essentially de-humanize women. The real disagreement turns out to be over a non-moral matter: whether marriage de-humanizes women. Despite appearances, you might actually *share* the speaker's moral outlook.

When someone makes a case for some moral position that you find counter-intuitive, reconstructing her or his reasoning as a moral syllogism is often really helpful: it is a quick method for locating the source of disagreement. Now, people often give their reasons in the form of **enthymemes**, i.e. syllogisms with an unstated premise. Sometimes a speaker will suppress a premise because she or he thinks it's too obvious to be said, or—more underhandedly—because the premise is highly questionable and would raise questions were it stated. Reconstructing enthymematic arguments requires charity, creativity, and trial-and-error. Let the form be your guide.

8.3 Moral Analogy

Moral syllogisms start with a general moral principle and then assess a particular item (usually an action) by claiming that it falls under that principle. Using moral syllogisms in our own reasoning, and when giving arguments to others, helps us thinking coherently and systematically about moral matters. Sometimes, though, you won't know which *general* moral principles should be brought to bear in a case, either because the case is unusual or unprecedented, or because you don't know what general moral principles your conversation partner accepts, or because you're just getting started in developing a systematic moral outlook. On such occasions, you're not in a position to argue from general moral principles to specific cases. Rather, you will have to use one specific case to guide your thinking about another specific case. In other words, you need to use an argument from analogy.

Human sexuality is one of the domains of life that has proven tricky to talk about in terms of general moral principles. This is not to say that one's views on human sexuality can't be part of a systematic moral outlook. Often, however, general moral principles that apply to human sexuality are just as much in dispute as the particular judgements that follow from those principles. To illustrate, suppose that Anna and Brian are arguing about same-sex marriage. Each provides the other with a moral syllogism, as follows:

Anna:

- P1. All sexual arrangements that are not ordered toward the loving, covenanted, and potentially procreative union of male and female are morally inappropriate.
- P2. Same-sex marriage is a sexual arrangement that is not ordered toward the loving, covenanted, and potentially procreative union of male and female.
- C. So, same-sex marriage is morally inappropriate.

Brian:

- P1. All sexual arrangements that are free, consensual, and harmless to other parties are morally appropriate.
- P2. Same-sex marriage is a sexual arrangement that is free, consensual, and harmless to other parties.

C. So, same-sex marriage is morally appropriate.

The trouble is that someone who rejects the conclusion of Anna's argument is likely also to reject premise 1 in that argument, and someone who rejects the conclusion of Brian's argument is likely also to reject premise 1. To the extent that Anna and Brian disagree about same-sex marriage, they are likely also to disagree about the general moral principles governing sexuality. (Note: though these arguments won't be *persuasive*, they are still *clarifying*, in the manner describes at the end of the previous section. Anna and Brian much better understand one another's thinking. That's an important kind of progress in discussion.)

But Anna and Brian are liable to agree about *some* things related to sexuality. If Anna wants to make her case about same-sex marriage, she needs to identify some particular sexual arrangement or arrangements that Brian already believes is morally inappropriate, and then argue that there is a strong analogy between that arrangement and same-sex marriage. Brian, likewise, needs to identify a sexual arrangement or arrangements that Anna already believes is morally appropriate. For example,

Anna:

P1. Polygamy and incest, though sought-after by a minority of the population, have been deemed deviant and disordered by mainstream society for centuries, and are morally inappropriate.
P2. Same-sex marriage, though sought-after by a minority of the population, has been deemed deviant and disordered by mainstream society for centuries.
C. So, same-sex marriage is morally inappropriate.

Brian:

P1. Heterosexual marriage is the uncoerced forming of a family unit, motivated by mutual love, and is morally appropriate.
P2. Same-sex marriage is the uncoerced forming of a family unit, motivated by mutual love.
C. So, same-sex marriage is morally appropriate.

Neither Anna nor Brian are likely to be immediately persuaded by the other's argument. Rather, their arguments are likely to spur further discussion about the morally relevant similarities and differences between same-sex marriage, on the one hand, and the primary analogues appealed to in each argument, on the other.

These arguments are **moral analogies.** Construed very broadly, a moral analogy is an inference from the moral status of one or more items to a conclusion about the moral status of a similar item. The form for moral analogy is as follows:

P1. x1, x2, x3, etc.… are A, B, C, etc., and M.
P2. y is A, B, C, etc.
C. So, y is M.

Moral analogies are straightforward substitution-instances of the form for argument from analogy, so the form for moral analogy is exactly the same—with one specification.

Unlike non-moral arguments from analogy, their target-feature is always a **moral property** M (in our examples, *being morally inappropriate* and *being morally appropriate*).

Of course, there are plenty of other domains of human life besides sexuality where moral analogies can be helpful. Whenever there are uncertainties or disagreements about the moral status of something, it is helpful to try to identify similar cases that already enjoy consensus. Now, our moral intuitions are often clearest when it comes to cases of extreme moral badness, e.g. slavery, genocide, child abuse, and other forms of egregious evil. It is very common for speakers to try to compare some social practice they disapprove of to one of these extreme cases. This is rarely productive, because there are almost always a lot of relevant differences between the clear case and the case in question.

Judicial reasoning is a context where moral syllogisms and moral analogies often support one another. Suppose a defendant has been caught trying to steal gasoline from a gas station. She pleads innocence on the grounds that her abusive boyfriend threatened violence if she did not steal the gasoline. Should the judge find her guilty? The prosecuting attorney points out that her action falls under the legal code that prohibits theft. The defense attorney, on the other hand, points out that her action falls under the legal code that exonerates those who commit crimes under severe duress. The judge has thus been presented with two moral syllogisms, one of which concludes that, prima facie, the judge should find the defendant guilty, and the other of which concludes, prima facie, the judge should find the defendant innocent. How should the judge proceed?

What judges actually do in such cases is appeal to *precedent*. That is, they look to past cases that are similar in relevant respects to the case at hand. If the weight of precedent falls on a particular side or the other—if, for example, defendants in similar cases have tended to be found innocent—then the judge can construct an argument from analogy, with defendants in these past cases as primary analogues, in order to 'break the tie' between the two prima facie verdicts. Of course, precedents may be mixed; different past cases may have different base features in common with ongoing cases; and past judgements can be determined to have been in error. Consequently, reasoning in courts of law are often very complicated and extremely interesting. Transcripts of hearings in the U.S. Supreme Court, for example, often read as a battery of moral analogies, each of which is subjected to intense scrutiny by everyone in the room. Just as science journalism tends to omit crucial, qualifying details of causal studies, so legal and political journalism tends to omit much of the fascinating, nuanced reasoning that goes into high-profile judicial decisions.

8.4 Moral Generalization

Moral syllogisms start from general principles and reach conclusions about particular cases. Moral analogies start from particular cases and reach conclusions about other cases. **Moral generalizations** start from particular cases and then generalize to principles.

Consider the following exchange between Claude, a socialist, and Olivia, a libertarian.

Claude: The federal government is complicit in the absurd wealth inequality in this country. Justice demands that we increase federal income taxes on the wealthy and then use this new revenue to expand our social welfare programs.

Olivia: I disagree. Yes, wealth inequality is a social problem. And the rich *should* contribute their wealth toward programs that close the gap. But it is not the government's job to make sure they do so.

Claude: Why ever not?

Olivia: Here's why. Imposing one particular code of moral conduct on society is inappropriate in a free society. Legally coercing rich people to give money to social programs is an instance of imposing one particular code of moral conduct on society. So, legally coercing rich people to give money to social programs is inappropriate in a free society.

Claude: You just invoked the principle that the government shouldn't be imposing a moral code of conduct on society. But where is that written? Why should we think that principle is true?

Olivia: Well, let's think about some particular cases. Suppose the government made it a crime to eat meat, because the people currently in power believed that vegetarianism is morally required. Do you think such a law would be appropriate in a free society?

Claude: I do not.

Olivia: Ok, now suppose a majority in Congress believed that gambling is a sin, and so they passed a law making it a crime to play Poker, or to have an office March Madness pool. Do you think that would be ok?

Claude: No, that would be a case of the government sticking its nose where it doesn't belong.

Olivia: Oh, and remember when our legislators became convinced that alcohol was a moral scourge on society, and proceeded to make it illegal to buy, sell, or drink alcohol? You aren't in favor of bringing Prohibition back, are you?

Claude: Of course not.

Olivia: The three cases I just described are all cases of legislation that imposes a particular moral code on society, and you agreed that all three types of legislation are inappropriate in a free society. That's why we should accept the moral principle I began with—that imposing one particular code of moral conduct on society is inappropriate in a free society.

Notice what has happened. Olivia provided an argument against Claude's view in the form of a moral syllogism, which we can reconstruct as follows:

P1. Imposing one particular code of moral conduct on society is inappropriate in a free society.
P2. Legally coercing rich people to give money to social programs is an instance of imposing one particular code of moral conduct on society.
C. Legally coercing rich people to give money to social programs is inappropriate in a free society.

Claude then requested a rationale for the general moral principle expressed in the first premise. (Claude could just as easily have requested a rationale for the second, non-moral premise, but that's not the line of critique he pursued.) Olivia accepted the challenge by providing illustrative cases—cases of illegitimate impositions by the government of a moral code—and then generalizing from them. Here's a reconstruction of her argument:

P1. Legal prohibitions on meat-eating, gambling, and drinking alcohol are cases of the government imposing one particular moral code of conduct on society.
P2. Legal prohibitions on meat-eating, gambling, and drinking alcohol are all inappropriate in a free society.
C. So, all cases of imposing one particular moral code of conduct on society are inappropriate in a free society.

Olivia's inference is an inductive generalization. But it is an inductive generalization of particular sort. For example, it involves describing little scenarios or rehearsing little narratives, and these scenarios and narratives can be fictional. (Prohibition happened in the 1920s and '30s, but to date the U.S. government has not prohibited meat-eating or gambling.) Though we *could* use the form for inductive generalizations we learned in chapter 4 above to reconstruct moral generalizations, a slightly modified form is a little more user-friendly:

P1. x1, x2, x3, etc., are cases of As.
P2. x1, x2, x3, etc. are all M.
C. So, all As are M.

The first premise states that one or more **cases** (x1, x2, x3, etc.) are cases of the **population** A, where this 'population' is a set of cases that all have some feature in common. (In our example, the population is *impositions of one particular moral code of conduct on society*). The second premise states that the case or cases all have some **moral property** M (in our example, *being appropriate in a free society*). Finally, the conclusion states that *all* members of the population—not just the cases mentioned in the first premise—have the value-property.

Mirroring what we have learned about other inductive generalizations, the strength of moral generalization turns on whether the cases invoked are representative of the 'population'. What this means is that the cases should be sufficiently numerous and diverse that *the only morally relevant commonality among them is the feature that characterizes the population*. Returning to our example: Olivia is claiming that the morally relevant commonality among the cases she mentions is that they all involve impositions of a particular code of moral conduct on the society. Claude can grant that they *are* members of this 'population', but insist that the generalization is unwarranted—if he can identify some other, or less general, morally relevant feature that the cases have in common. For example, maybe what makes prohibitions on meat-eating, gambling, and drinking alcohol inappropriate is they are all *prohibitions* (as opposed to other sorts of uses of the law), or that all three involve activities that are all 'private' rather than public, or that these activities don't directly harm anyone, or that they don't involve commerce in any way, or that they involve what one does with one's *net* income (i.e., income after taxes), and so on. We can imagine Claude responding to Olivia in the following way:

> Claude: I don't buy the argument. We were initially talking about government-run income redistribution. You argued against it on the grounds that it's always inappropriate for the government to impose a moral code of conduct on the citizenry, and your argument for *that* claim was a generalization from three cases of the government illegitimately imposing its moral will on the citizenry. But each of your three cases involves the government's *prohibiting private activities that don't harm anybody*. Not all cases of the government's imposing a moral code on society involve doing that. How about when a government creates a new institution, for example Child Protective Services? Surely the government is warranted in "imposing" its conviction that child abuse is not ok. So, I don't think your argument is very strong.

At this point Olivia could respond by denying that her cases are special in the way Claude alleges, or she could expand her inventory of cases. The lesson is that it is tricky to tell whether a set of cases really is representative of some broader category. Careful, sustained reflection is required.

Two of Olivia's three cases are entirely made up. This is fine. When we're engaged in moral deliberation, it works just as well to appeal to hypothetical actions that *would* be good or bad were someone to engage in them, as to point to actual actions that *were* good and bad when someone *did* engage in them. If it weren't true that fictional cases could reveal moral truth, nobody could acquire moral knowledge from reading stories. But storytelling is among the oldest and most widespread means of communicating moral truth. (This use of storytelling in moral deliberation is part of what makes moral deliberation engaging, even fun.) Exactly *how* stories communicate moral truth is somewhat mysterious, and a matter of debate among scholars. Certainly *one* way they do so is by providing cases that can be used in moral generalizations. That is, by showing us how things can go morally well or badly for particular characters in concrete scenarios, they allow us to extract general moral principles about how things can go well or badly for anyone in similar circumstances.

8.5 Coda: From Giving Arguments to Having Productive Discussions

Our study of deductive and inductive arguments in this course has treated arguments primarily as *monologues*: premises given by a speaker in support of a conclusion. We have discussed how to recognize these arguments, categorize them, reconstruct them, and evaluate them.

But the term 'argument' is also used to refer to arguments *as dialogues*: reasons exchanged and responded to, in an ongoing back-back-forth between two or more parties. Arguments in the first sense are *given* by one person; arguments in the second sense are *had* by two or more people. The dialogue between Claude and Olivia in the previous section is an example of a snippet of an argument in this sense. *Having* an argument in the sense of a dialogue involves the *giving* of arguments in the sense of a monologue, but it involves much more than that. An argument in the dialogical sense does not follow any step-wise path from premises to conclusion, but meanders through whatever territory into which its participants take it. No single speaker is in control of it. If it is done right, every participant walks away having discovered something.

G.K. Chesterton wrote that "People generally quarrel because they cannot argue." An argument is *not* a mere heated disagreement between two people. (Productive arguments can get heated, but that's not what makes them productive.) Nor is it the opposite of a quarrel, a mere *sharing* and respectful acknowledgement of differing ideas. In an argument, my ideas are not 'safe'; anything I say is subjected to scrutiny by the other participants. But the scrutiny is *rational* scrutiny, and not everyone who wishes to engage in argumentation has developed the tools necessary to do that well. People quarrel because they *cannot* argue; they lack the rational tools necessary. But not you, not anymore—now that you have taken this course.

To avoid confusing arguments-in-the-dialogue-sense with quarrels, let's call these reason-governed conversations "productive discussions." What are the essential elements of a productive discussion? There are at least four. First, someone needs to give an argument (in the monologue-sense) about something, and it needs to be recognized and received *as* an argument, i.e., as a *rational* case for a claim (rather than a mere emotional appeal, hostile provocation, personal manifesto, etc.). This means that both parties need to know a little bit about what arguments (in the monologue-sense) are and what can make them better or worse. Sometimes this first stage of a productive discussion can take a long time, because the initial argument might need some refinement as it gets reconstructed. One's initial pass at making an argument often includes a lot of obscurity and confusion that can get cleared up as one's conversation partner asks clarificatory questions.

Second, the original argument needs to be subjected to rational assessment. That is, the other participants in the conversation need to point out potential shortcomings (maybe it's invalid, unsound, inductively weak, uncogent, or incomplete) and its virtues (maybe it's valid, sound, strong, cogent, or complete—or maybe it could be made so with some tweaking). If nobody in the conversation knows how to subject an argument to rational assessment—or even if some do and some don't—the discussion will quickly go off the rails. From here, a new cycle of giving and evaluating reasons can begin.

We have been working in our course on skills necessary to put these first two elements of a productive discussion in place. But a logic course cannot help with the last two. The third element is *time*. Productive discussions are likely to take place over the course of afternoons or whole weekends. More often than not, walking is involved, sometimes over great distances. They rarely happen on social media, and not a single one has ever happened on Twitter.

The fourth element is *intellectual virtue*. All participants in the discussion must place *knowledge of the truth* above any other motivation (such as loyalty to one's party, a desire to appear smart, the need to feel heard, resentment and indignation directed at others in the conversation, etc.). Intellectual virtue is not the natural state of human beings, and even the best of disputants sometimes act out of intellectual vice. (If you notice that you have done so, the thing to do is to pause the conversation, apologize, and move on with renewed resolve.) Ultimately, one cannot be a good discussion partner without being a virtuous human being. Only those of good character can use the rational tools of logic in productive ways.

Chapter 8 Homework Exercises

Instructions: Each of the following includes either a moral syllogism, a moral analogy, or a moral generalization. State which, and then reconstruct it as a substitution-instance of the relevant form discussed in class.

Example
Suppose every tourist who visited the Taj Mahal threw a penny at it. No single throw makes a noticeable difference; but over time, the building begins to chip and crumble. Would it be ok to say, "*My* throw won't make any noticeable difference," and proceed to throw the penny? No. It would be morally impermissible for you to be part of the gradual, destructive process. Likewise, each time you drive a car, you are contributing to the gradual destruction of the environment. Saying "*My* trip won't make any noticeable difference" and proceeding to drive is just as morally impermissible.

Answer:
Moral Analogy

P1. Throwing a penny at the Taj Mahal is contributing to a gradual, destructive process, and is morally impermissible.
P2. Driving a car is contributing to a gradual, destructive process.
C. So, driving a car is morally impermissible.

1. 16-year-olds should be allowed to vote. Any citizen who can engage in the democratic process via rational debate should be allowed to vote, and 16-year-olds certainly can.

2. 16-year-olds should be allowed to vote. I teach at a high school. Let me tell you about some of the students I teach. There's Cindy, who has been canvassing for local Republican candidates since she was in 8th grade. There's Doris, who petitioned the school board for expanded music and art requirements. There's Diego, who has organized several multi-cultural events for students. All of these citizens are 16 years old.

3. 16-year-olds should be allowed to vote. After all, we allow 18-year-olds to vote. But 16-year-olds are similar in all relevant respects to 18-year-olds: they can engage in critical thinking, and they have been educated on how government works.

4. If someone was trying to rape your spouse, and in an attempt to intervene you wind up killing the attacker, you would not be culpable for the attacker's death. So, you're not guilty of murder if you kill someone in an attempt to save someone from a perceived threat.

5. If you cause a stampede by yelling 'fire' in a crowded theater, you should be charged with reckless endangerment. Likewise, when cable news organizations drum up fear via highly biased reporting, they should not be able to use the fifth amendment as a protection against being charged with libel.

6. You shouldn't kick your sister. Can't you see it hurts her when you do that? You shouldn't do things that hurt people.

7. Even if you're not religious, you should lower your voice when you enter a church. After all, you should lower your voice when you enter a library, even if you don't care about books or reading. In both cases, there's a special purpose to the space that requires quiet.

8. A shipowner was about to send to sea an emigrant-ship. He knew that she was old, and not overwell built at the first; that she had seen many seas and climes, and often had needed repairs. Doubts had been suggested to him that possibly she was not seaworthy. These doubts preyed upon his mind, and made him unhappy; he thought that perhaps he ought to have her thoroughly overhauled and refitted, even though this should put him to great expense. Before the ship sailed, however, he succeeded in overcoming these melancholy reflections. He said to himself that she had gone safely through so many voyages and weathered so many storms that it was idle to suppose she would not come safely home from this trip also. He would put his trust in Providence, which could hardly fail to protect all these unhappy families that were leaving their fatherland to seek for better times elsewhere. He would dismiss from his mind all ungenerous suspicions about the honesty of builders and contractors. In such ways he acquired a sincere and comfortable conviction that his vessel was thoroughly safe and seaworthy; he watched her departure with a light heart, and benevolent wishes for the success of the exiles in

their strange new home that was to be; and he got his insurance-money when she went down in mid-ocean and told no tales.

What shall we say of him? Surely this, that he was verily guilty of the death of those men. It is admitted that he did sincerely believe in the soundness of his ship; but the sincerity of his conviction can in no wise help him, because he had no right to believe on such evidence as was before him….

There was once an island in which some of the inhabitants professed a religion teaching neither the doctrine of original sin nor that of eternal punishment. A suspicion got abroad that the professors of this religion had made use of unfair means to get their doctrines taught to children. They were accused of wresting the laws of their country in such a way as to remove children from the care of their natural and legal guardians; and even of stealing them away and keeping them concealed from their friends and relations. A certain number of men formed themselves into a society for the purpose of agitating the public about this matter. They published grave accusations against individual citizens of the highest position and character, and did all in their power to injure these citizens in their exercise of their professions. So great was the noise they made, that a Commission was appointed to investigate the facts; but after the Commission had carefully inquired into all the evidence that could be got, it appeared that the accused were innocent. Not only had they been accused on insufficient evidence, but the evidence of their innocence was such as the agitators might easily have obtained, if they had attempted a fair inquiry. After these disclosures the inhabitants of that country looked upon the members of the agitating society, not only as persons whose judgment was to be distrusted, but also as no longer to be counted honourable men. For although they had sincerely and conscientiously believed in the charges they had made, *yet they had no right to believe on such evidence as was before them*….To sum up: it is wrong always, everywhere, and for anyone, to believe anything upon insufficient evidence.

(William Clifford, "The Ethics of Belief")