

## Revised Fundamental Methods of Logic



# Revised Fundamental Methods of Logic

*MATTHEW KNACHEL AND SEAN GOULD*



*Revised Fundamental Methods of Logic by Matthew Knachel and Sean Gould is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, except where otherwise noted.*

# Contents

Revised Fundamental Methods of Logic	vii
Publication Information	viii
Table of Contents	ix
Preface	xiv
 Part I. Main Body	
 1. Chapter 1	1
Matthew Knachel	
2. Chapter 2	21
Matthew Knachel	
3. Chapter 3	47
Matthew Knachel	
4. Chapter 4	86
Matthew Knachel	
5. Chapter 5	115
<i>Deductive Logic III: Natural Proofs in Propositional Logic</i>	
Sean Gould	
6. Chapter 6	145
<i>Deductive Logic IV: Quantification and Predicate Logic</i>	
Sean Gould	
7. Chapter 7	165
Matthew Knachel	
8. Chapter 8	182
Matthew Knachel	



by Matthew Knachel  
Revision by Sean Gould

Revised Fundamental Methods of Logic is licensed under Creative Commons Attribution 4.0 International License. CC-BY-NC.

Cover Art by Sean Gould and Barry Biechner

The original version of this book is published by the UW-Milwaukee Library Digital Commons and is available free of charge at this site: [http://dc.uwm.edu/phil\\_facbooks/1](http://dc.uwm.edu/phil_facbooks/1)



This work is licensed under the Creative Commons Attribution 4.0 International License.

For details, visit this site: <https://creativecommons.org/licenses/by/4.0/legalcode>

Cover Art on original version by Dan Williams

Recommended citation for original version: Knachel, Matthew, "Fundamental Methods of Logic" (2017).

Philosophy Faculty Books. 1. [http://dc.uwm.edu/phil\\_facbooks/1](http://dc.uwm.edu/phil_facbooks/1)

ISBN: 978-0-9961502-2-4

*For my girls, Rose and Alice – Matthew Knachel*



## Preface

## Chapter 1 – The Basics of Logical Analysis

- I. What is Logic?
- II. Basic Notions: Propositions and Arguments
- III. Recognizing and Explicating Arguments
  - Paraphrasing
  - Enthymemes: Tacit Propositions
  - Arguments vs. Explanations
- IV. Deductive and Inductive Arguments
  - Deductive Arguments
  - Inductive Arguments
- V. Diagramming Arguments
  - Independent Premises
  - Intermediate Premises
  - Joint Premises

## Chapter 2 – Informal Logical Fallacies

- I. Logical Fallacies: Formal and Informal 29
- II. Fallacies of Distraction 31
  - Appeal to Emotion (*Argumentum ad Populum*) 31
  - Appeal to Force (*Argumentum ad Baculum*) 32
  - Straw Man 33
  - Red Herring 34
  - Argumentum ad Hominem* 36
- III. Fallacies of Weak Induction 30
  - Argument from Ignorance (*Argumentum ad Ignorantiam*) 40
  - Appeal to Inappropriate Authority 42
  - Post hoc ergo propter hoc* 43
  - Slippery Slope 44
  - Hasty Generalization 45
- IV. Fallacies of Illicit Presumption 46
  - Accident 46
  - Begging the Question (*Petitio Principii*) 48
  - Loaded Questions 49
  - False Choice 51
  - Composition 52
  - Division 53
- V. Fallacies of Linguistic Emphasis 53
  - Accent 54

Quoting out of Context	55
Equivocation	57
Manipulative Framing	59

## Chapter 3 – Deductive Logic I: Aristotelian Logic 68

I. Deductive Logics	68
II. Classes and Categorical Propositions	69
The Four Types of Categorical Proposition	71
Universal Affirmative (A)	71
Universal Negative (E)	74
Particular Affirmative (I)	74
Particular Negative (O)	76
A Note on Terminology	76
Standard Form for Sentences Expressing Categorical Propositions	77
III. The Square of Opposition	79
Contradictories	80
Contraries	81
Subcontraries	81
Subalterns	82
Inferences	83
IV. Operations on Categorical Sentences	84
Conversion	84
Obversion	86
Contraposition	87
Inferences	91
V. Problems with the Square of Opposition	96
Existential Import	97
Problems for the Square	97
Solution?	98
Boolean Solution	99
VI. Categorical Syllogisms	102
Logical Form	103
The Venn Diagram Test for Validity	104

## Chapter 4 – Deductive Logic II: Propositional Logic 117

I. Why Another Method of Deductive Logic?	117
II. Syntax of PL	120
Conjunctions	121
Disjunctions	121
Negations	121
Conditionals	122
Biconditionals	122

Punctuation – Parentheses	123
II. Semantics of SL	126
Negations (TILDE)	127
Conjunctions (DOT)	128
Disjunctions (WEDGE)	128
Biconditionals (TRIPLE-BAR)	129
Conditionals (HORSESHOE)	130
III. Computing Truth-Values of Compound SL Sentences	133
IV. Translating from English into SL	
Tilde, Dot, Wedge	
Horseshoe and Triple-Bar	
V. Testing for Validity in SL	
Logical Argument Form in SL	
The Truth Table Test for Validity	

## Chapter 5 – Deductive Logic III: Natural Proofs in Propositional Logic

I. Replacement and Rules of Inference	
II. Rules of Inference	
Modus Ponens (M.P.)	
Modus Tollens (M.T.)	
Hypothetical Syllogism (H.S.)	
Disjunctive Syllogism (D.S.)	
Constructive Dilemma (C.D.)	
Simplification (Simp.)	
Conjunction (Conj.)	
Absorption (Abs.)	
Addition (Add.)	
Conclusion	
III. Replacement Rules	
Double Negation (D.N.)	
Commutation (Com.)	
Association (Assoc.)	
Material Implication (Impl.)	
De Morgan's Theorems (De M.)	
Distribution (Dist.)	
Material Equivalence (Equiv.)	
Exportation (Exp.)	
Tautology (Taut.)	
IV. Constructing Formal Proofs Using Rules of Inference and Replacement	
V. Conditional Proofs	
The Anatomy of a Conditional Proof	
Using Multiple Conditional Proofs	
Conditional Proofs for Tautologies	
VI. Indirect Proofs	
Proving Tautologies with Indirect Proofs	

## Chapter 6 – Deductive Logic IV: Quantification and Predicate Logic

- I. Affirmative Singular Predicates
- II. Existential and Universal Quantifiers
- III. Subject-Predicate Propositions
- IV. Instantiation, Generalization, and the Quantifier Negation Rule
- V. Formal Proofs of Validity for Aristotelian Syllogisms
  - Universal Instantiation
  - Universal Generalization
  - Existential Generalization
  - Existential Instantiation
- VI. Asyllogistic Inference
- VII. Proving Validity and Invalidity with Predicate Logic

## Chapter 7 – Inductive Logic I: Analogical and Causal Arguments

- I. Inductive Logics
- II. Arguments from Analogy
  - The Form of Analogical Arguments
  - The Evaluation of Analogical Arguments
    - Number of Analogues
    - Variety of Analogues
    - Number of Similarities
    - Relevance of Similarities and Differences
    - Modesty / Ambition of the Conclusion
  - Refutation by Analogy
- III. Causal Reasoning
  - The Meaning(s) of 'Cause'
  - Mill's Methods
    - Method of Agreement
    - Method of Difference
    - Joint Method of Agreement and Difference
    - Method of Residues
    - Method of Concomitant Variation
  - The Difficulty of Isolating Causes

## Chapter 6 – Inductive Logic II: Probability and Statistics

- I. The Probability Calculus
  - Conjunctive Occurrences
  - Disjunctive Occurrences
- II. Probability and Decision-Making: Value and Utility

III. Probability and Belief: Bayesian Reasoning	
IV. Basic Statistical Concepts and Techniques	
Averages: Mean vs. Median	
Normal Distributions: Standard Deviation, Confidence Intervals	
Statistical Inference: Hypothesis Testing	
Statistical Inference: Sampling	
V. How to Lie with Statistics	
Impressive Numbers without Context	
Misunderstanding Error	
Tricky Percentages	
The Base-Rate Fallacy	
Lying with Pictures	

## Preface to This Version,

Matthew Knachel's *Fundamental Methods of Logic* is a wonderfully accessible text for use in Introduction to Logic courses. The text benefits from Knachel's experience teaching the subject in many ways. The text is rich with examples. The subject matter and content are presented with sufficient thoroughness as to ensure students can achieve a substantive familiarity with the topics while not being overwhelmed by complexities. Fundamental skills are given significant attention enabling students to create and retain a strong foundation for the application of logical analysis to matters beyond the course itself.

*Fundamental Methods of Logic* is published as an open educational resource (OER) with an open Creative Commons license. Knachel gave this book to any student or teacher to use. Not only does the book have an open copyright for use, but Knachel also opened the book to future edits and revisions. This edition is an example of Knachel's generosity.

This new version comes with two new chapters. A new, fifth chapter now covers formal proofs in propositional logic. Chapter 5 is written as a stand-alone chapter with only non-essential references and cross-citations to the previous chapters. The chapter explains how to construct formal proofs using 19 rules of inference and replacement, as well as conditional and indirect proofs.

A new, sixth chapter introduces predicate logic and quantification. This chapter is also written as a potentially free-standing chapter, with minor references to parallels that occur in Chapter 3 which covers Aristotelian syllogism. Chapter 6 introduces quantification, rules of universal and existential generalization and instantiation, as well as formal proofs using predicate logic. Conditional and Indirect proofs are re-introduced in this new context. The chapter also covers the use of Truth Tables to prove invalidity using models with  $1 - n$  objects.

Both new chapters try, as much as is possible, to match the tone, style, and formatting of the pre-existing chapters so that readers can work through the complete, new version with continuity and congruity.

This new version also includes several non-substantive edits to Knachel's original work. A few clarifying sentences have been added at various points throughout any given chapter. Some of the tangential footnotes that were lost as the text transferred between file formats during editing were not replaced out of the sake of concision. At other places, some of the potentially more opinionated and expressive examples were neutralized.

Most dramatically, however, this new version avoids the previous version's use of examples that draw upon U.S. politics. Knachel's repeated references to the 2016 presidential election, Donald Trump, and Hillary Clinton have been replaced with difference examples. I think that while Knachel's use of contemporary matters can help engage students. However, in different localities and contexts such examples can also over-engage them and lead to distracting tangential debates regarding politics, rather than logic. For better or worse, it is unclear the extent to which politics and logic mix.

Thank you, Matthew Knachel, for sharing a solid open education resource with the educational community. I could not have written the additional chapters 5 and 6 without using Copi, Cohen, and Rodych's *Introduction to Logic* as a guide. Just as Knachel gives credit to his previous logic instructors, I would like to give credit to Erin Anchustegui and Mary Tiles. I would also like to acknowledge Liza Long, Greg Wilson, Anchustegui, and Jean-Louise Zancanella for their support revising and working with this book and OER materials in general. Financial support for my time was provided through funds from the College of Western Idaho's OER stipend program and an OPAL fellowship from the Idaho State Board of Education.

Sean Gould

## Knachel's Original Preface.

There's an ancient view, still widely held, that what makes human beings special—what distinguishes us from the “beasts of the field”—is that we are rational. What does rationality consist in? That's a vexed question, but one possible response goes roughly like this: we manifest our rationality by engaging in activities that involve *reasoning*—making claims and backing them up with reasons, acting in accord with reasons and beliefs, drawing inferences from available evidence, and so on.

This reasoning activity can be done well, and it can be done badly—it can be done correctly or incorrectly. Logic is the discipline that aims to distinguish good reasoning from bad.

Since reasoning is central to all fields of study—indeed, since it's arguably central to being human—the tools developed in logic are universally applicable. Anyone can benefit from studying logic by becoming a more self-aware, skillful reasoner.

This covers a variety of topics at an introductory level. Chapter One introduces basic notions, such as arguments and explanations, validity and soundness, deductive and inductive reasoning; it also covers basic analytical techniques, such as distinguishing premises from conclusions and diagramming arguments. Chapter Two discusses informal logical fallacies. Chapters Three and Four concern deductive logic, introducing the basics of Aristotelian and Sentential Logic, respectively. Chapters Five and Six concern inductive logic. Chapter Five deals with analogical and causal reasoning, including a discussion of Mill's Methods. Chapter Six covers basic probability calculations, Bayesian inference, fundamental statistical concepts and techniques, and common statistical fallacies.

The text is suitable for a one-semester introductory logic or “critical thinking” course. The emphasis is on formal techniques and problem solving rather than analytical writing, though exercises of the latter sort could easily be incorporated.

A note on tone, style, and content. This book is written by an American teacher whose intended audience is American undergraduates; it is based on my lectures, developed over many years. Like the lectures, it assumes that some members of the intended audience lack an antecedent interest in the subject and may have trouble developing and maintaining enthusiasm to study it. It tries to compensate for this by adopting a casual style, using first- and second-person constructions, and by shamelessly trafficking in cultural references, lame jokes, and examples involving American current events. The result is a logic textbook with a somewhat unusual tone and a sometimes narrow cultural perspective. Neither familiarity with the relevant cultural references, nor amusement at the lame jokes, is a prerequisite for understanding the material, but I thought it prudent to offer an *apologia* at the outset. *Caveat lector*.

An acknowledgment of debts. The following books have influenced my teaching, and hence the present work: Virginia Klenk's *Understanding Symbolic Logic*, John Norton's *How Science Works*, Ian Hacking's *Introduction to Probability and Inductive Logic*, Darrell Huff's *How to Lie with Statistics*, and Irving Copi and Carl Cohen's *Introduction to Logic*. The influence of those last two books is particularly profound, as I note throughout this text. I am indebted to all my logic teachers over the years: Kurt Mosser, Michael Liston, Mark Kaplan, Richard Tierney, Steve Leeds, Joan Weiner, Ken Manders, Mark Wilson, and Nuel Belnap. Thanks to J.S. Holbrook for sending me examples of fallacies. For extensive logistical support, I'm indebted to Kristin Miller Woodward; I also thank her for arranging financial support through the UW-Milwaukee Library and Center for Excellence in Teaching and Learning, who have undertaken a project to encourage the development and adoption of open textbooks. My logic students over the years also deserve acknowledgment, especially those who have recently served as guinea pigs, learning from earlier drafts of this book. Without student feedback, there would be no book. Finally, and most importantly, I could not have completed this project without my wife Maggie's constant support and forbearance.





# I.

MATTHEW KNACHEL

## The Basics of Logical Analysis

### I. What Is Logic?

In Logic, the object of study is *reasoning*. This is an activity that humans engage in—when we make claims and back them up with reasons, or when we make inferences about what follows from a set of statements.

Like many human activities, reasoning can be done well, or it can be done badly. The goal of logic is to distinguish good reasoning from bad. Good reasoning is not necessarily effective reasoning; in fact, as we shall see, bad reasoning is pervasive and often extremely effective—in the sense that people are often persuaded by it. In Logic, the standard of goodness is not effectiveness in the sense of persuasiveness, but rather *correctness* reliably established by logical rules. In logic, we study the rules and techniques that allow us to distinguish good, correct reasoning from bad, incorrect reasoning.

Since there is a variety of different types of reasoning, since it's possible to develop various methods for evaluating each of those types, and since there are different views on what constitutes correct reasoning, there are many approaches to the logical enterprise. We talk of logic, but also of *logics*. A logic is just a set of rules and techniques for distinguishing good reasoning from bad. There are many logics; the purpose of this book is to give an overview of some of the most basic ones.

So, the object of study in logic is correct reasoning, with the goal of distinguishing the good from the bad. Logic is not the study of how people think. It is important to note that this approach sets logic apart from an alternative way of studying human reasoning, one more proper to a different discipline: psychology. It is possible to study human reasoning in a merely *descriptive* mode: to identify common patterns of reasoning

and explore their psychological causes, for example. This is not logic. Logic takes up reasoning in a *prescriptive* mode: it tells how we *ought* to reason, not merely how we in fact typically do.<sup>1</sup>

### II. Basic Notions: Propositions and Arguments

Reasoning involves claims or statements—making them and backing them up with reasons, drawing out their consequences. *Propositions* are the things we claim, state, assert.

Propositions are the kinds of things that can be true or false. They are expressed by declarative sentences. 'This book is boring' is a declarative sentence; it expresses the proposition that this book is boring, which is (arguably) true (at least

1. Psychologists have determined, for example, that most people are subject to what is called "confirmation bias"—a tendency to seek out information to confirm one's pre-existing beliefs, and ignore contradictory evidence. There are lots of studies on this effect, including even brain-scans of people engaged in evaluating evidence. All of this is very interesting, but it's psychology, not logic; it's a mere descriptive study of reasoning. From a logical, prescriptive point of view, we simply say that people should try to avoid confirmation bias, because it can lead to bad reasoning.

so far—but it's only the first page; wait until later, when things get exciting! You won't believe the cliffhanger at the end of Chapter 3. Mindblowing.).

We distinguish propositions from the sentences that express them because a single proposition can be expressed by different sentences. "It's raining" and "Es regnet" both express the proposition that it is raining; one sentence does it in English, the other in German. Also, "Clyde loves Irene" and "Irene is loved by Clyde" both express the same proposition.

Other kinds of sentences do not express propositions. Imperative sentences issue commands: 'Sit down and shut up' is an imperative sentence; it doesn't make a claim, express something that might be true or false; either it's obeyed or it isn't. Interrogative sentences ask questions: 'Who will win the World Cup this year?' is an interrogative sentence; it does not assert anything that might be true or false either.

Only declarative sentences express propositions, and so they are the only kinds of sentences we will deal with at this stage of the study of logic. (More advanced logics have been developed to deal with imperatives and questions, but we won't look at those in an introductory textbook.)

The fundamental unit of reasoning is the *argument*. In logic, by 'argument' we don't mean a disagreement, a shouting match; rather, we define the term precisely:

Argument = a set of propositions, one of which, the conclusion, is (supposed to be) supported by the others, the premises.

If we're reasoning by making claims and backing them up with reasons, then the claim that's being backed up is the conclusion of an argument; the reasons given to support it are the argument's premises. If we're reasoning by drawing an inference from a set of statements, then the inference we draw is the conclusion of an argument, and the statements from which it is drawn are the premises.

We include the parenthetical hedge—"supposed to be"—in the definition to make room for bad arguments. Remember, in Logic, we're evaluating reasoning. Arguments can be good or bad, logically correct or incorrect. A bad argument, very roughly speaking, is one where the premises fail to support the conclusion; a good argument's premises actually do support the conclusion.

To support the conclusion means, again very roughly, to give one good reason for believing it. Arguments are meant to preserve the truth of its premises as we reason toward a conclusion, thereby helping us accurately state things about the world. This highlights the *rhetorical purpose* of arguments: we use arguments when we're disputing controversial issues; they aim to persuade people, to convince them to believe their conclusion. As we said, in logic, we don't judge arguments based on whether or not they succeed in this goal—there are logically bad arguments that are nevertheless quite persuasive. Rather, the logical enterprise is to identify the kinds of reasons that *ought* to be persuasive (even if they sometimes aren't). We want to know which arguments ensure that true or likely true premises lead to true or likely true conclusions and which arguments lead us into error.

As this book proceeds, we will begin presenting arguments in a way that individual symbols are used to represent entire propositions. Part of this is so that we needn't write the entire proposition out every time we use it. Another reason is that the study of Logic is in large part the study of the connections and steps we make between propositions as we reason from premises to a conclusion.<sup>2</sup> Thus, we might say, "for the time being, let R stand for "it is raining," and let W stand for "it is wet outside." Then we might say if R then W. Learning how to think and talk about *reasoning* itself, regardless of whatever propositional content we are dealing with is a goal of this book.

2. Reasoning in the sense of drawing inferences from a set of statements is a special case of this persuasive activity. When we draw out reasonable conclusions from given information, we're convincing ourselves that we have good reasons to believe them.

### III. Recognizing and Explicating Arguments

Before we get down to the business of evaluating arguments—deciding whether they’re good or bad—we need to develop some preliminary analytical skills. The first of these is, simply, the ability to recognize arguments when we see them, and to figure out what the conclusion is (and what the premises are).

What we want to learn first is how to *explicate* arguments. This involves writing down a bunch of declarative sentences that express the propositions in the argument, and clearly marking which of these sentences expresses the conclusion. Let’s start with a simple example. Here’s an argument:

You really shouldn’t eat at McDonald’s. Why? First of all, they pay their workers very low wages. Second, the animals that go into their products are raised in deplorable, inhumane conditions. Third, the food is really bad for you. Finally, the burgers have poop in them.<sup>3</sup>

The passage is clearly argumentative: its purpose is to convince you of something, namely, that you shouldn’t eat at McDonald’s. That’s the conclusion of the argument. The other claims are all reasons for believing the conclusion—reasons for not eating at McDonald’s. Those are the premises.

To *explicate* the argument is simply to clearly identify the premises and the conclusion, by writing down declarative sentences that express them. We would explicate the McDonald’s argument like this:

McDonald’s pays its workers very low wages.

The animals that go into their products are raised in deplorable, inhumane conditions.

McDonald’s food is really bad for you.

Their burgers have poop in them.

/∴ You shouldn’t eat at McDonald’s.

We separate the conclusion from the premises with a special symbol in front of the conclusion, /∴, which can be read as “therefore.”

Speaking of ‘therefore’, it’s one of the words to look out for when identifying and explicating arguments. Along with words like ‘consequently’ and ‘thus’, and phrases like ‘it follows that’ and ‘which implies that’, it indicates the presence of the conclusion of an argument. Similarly, words like ‘because’, ‘since’, and ‘for’ indicate the presence of premises.

We should also note that it is possible for a single sentence to express more than one proposition. If we added this sentence to our argument—‘McDonald’s advertising targets children to try to create lifetime addicts to their high-calorie foods, and their expansion into global markets has disrupted native food distribution systems, harming family farmers’—we would write down two separate declarative sentences in our explication, expressing the two propositions asserted in the sentence—about children and international farmers, respectively. Indeed, it’s possible for a single sentence to express an entire argument. ‘You shouldn’t eat at McDonald’s because they’re a bad corporate actor’ gives you a conclusion and a premise at once. An explication would merely separate them.

3. Consumer Reports conducted a study in 2015, in which they tested 458 pounds of ground beef, purchased from 103 different stores in 26 different cities; all of the 458 pounds were contaminated with fecal matter. This is because most commercial ground beef is produced at facilities that process thousands of animals, and do it very quickly. The quickness ensures that sometimes—rarely, but sometimes—a knifecut goes astray and the gastrointestinal tract is nicked, releasing poop. It gets cleaned up, but again, things are moving fast, so they don’t quite get all the poop. Now you’ve got one carcass—again, out of hundreds or thousands—contaminated with feces. But they make ground beef in a huge vat, with meat from all those carcasses mixed together. So even one accident like this contaminates the whole batch. So yeah, those burgers—basically all burgers, unless you’re grinding your own meat or sourcing your beef from a local farm—have poop in them. Not much, but it’s there. Of course, it won’t make you sick as long as you cook it right: 160° F is enough to kill the poop-bacteria (E-coli, etc.).

## Paraphrasing

The argument about McDonald's was an easy case. It didn't have a word like 'therefore' to tip us off to the presence of the conclusion, but it was pretty clear what the conclusion was anyway. All we had to do was ask ourselves, "What is this person trying to prove?" The answer to that question is the conclusion of the argument.

Another way the McDonald's argument was easy: all of the sentences were declarative sentences, so when we explicated the argument, all we had to do was write them down. But sometimes argumentative passages aren't so cooperative. Sometimes they contain non-declarative sentences. Recall, arguments are sets of propositions, and only declarative sentences express propositions; so if an argumentative passage contains non-declarative sentences (questions, commands, etc.), we need to change their wording when we explicate the argument, turning them into declarative sentences that express a proposition. This is called *paraphrasing*.

Suppose, for example, that the McDonald's argument were exactly as originally presented, except the first sentence were imperative, not declarative:

Don't eat at McDonald's. Why? First of all, they pay their workers very low wages. Second, the animals that go into their products are raised in deplorable, inhumane conditions. Third, the food is really bad for you. Finally, the burgers have poop in them.

We just switched from 'You shouldn't eat at McDonald's' to 'Don't eat at McDonald's.' But it makes a difference. The first sentence is declarative; it makes a claim about how things are (morally, with respect to your obligations in some sense): you shouldn't do such-and-such. It's possible to disagree with the claim: Sure I should, and so should everybody else; their fries are delicious! 'Don't eat at McDonald's', on the other hand, is not like that. It's a command. It's possible to disobey it, but not to disagree with it; imperative sentences don't make claims about how things are, don't express propositions.

Still, the passage is clearly argumentative: the purpose remains to persuade the listener not to eat at McDonald's. We just have to be careful, when we explicate the argument, to paraphrase the first sentence—to change its wording so that it becomes a declarative, proposition-expressing sentence. 'You shouldn't eat at McDonald's' works just fine.

Let's consider a different example:

I can't believe anyone would support a \$15 per hour minimum wage. Don't they realize that it would lead to massive job losses? And the strain such a policy would put on small businesses could lead to an economic recession.

The passage is clearly argumentative: this person is engaged in a dispute about a controversial issue—the minimum wage—and is staking out a position and backing it up. What is that position? Apparently, this person opposes the idea of raising the minimum wage to \$15.

There are two problems we face in explicating this argument. First, one of the sentences in the passage—the second one—is non-declarative: it's an interrogative sentence, a question. Nevertheless, it's being used in this passage to express one of the person's reasons for opposing the minimum wage increase—that it would lead to job losses. So we need to paraphrase, transforming the interrogative into a declarative—something like 'A \$15 minimum wage would lead to massive job losses'.

The other problem is that the first sentence, while a perfectly respectable declarative sentence, can't be used as-is in our explication. For while it's clearly being used by to express this person's main point, the conclusion of his argument against the minimum wage increase, *it does so indirectly*. What the sentence literally and directly expresses is not a claim about the wisdom of the minimum wage increase, but rather a claim about the speaker's personal beliefs: 'I can't believe anyone would support a \$15 per hour minimum wage'. But that claim isn't the conclusion of the argument. The speaker isn't trying to convince people that he believes (or can't believe) a certain thing; he's trying to convince them to believe the same thing he believes, namely, that raising the minimum wage to \$15 is a bad idea. So, despite the first sentence being a declarative, we still have to paraphrase it. It expresses a proposition, but not the conclusion of the argument.

Our explication of the argument would look like this:

Increasing the minimum wage to \$15 per hour would lead to massive job losses.

The policy would put a strain on small businesses that might lead to a recession.  
/∴ Increasing the minimum wage to \$15 per hour is a bad idea.

### *Enthymemes: Tacit Propositions*

So sometimes, when we explicate an argument, we have to take what's present in the argumentative passage and change it slightly, so that all of the sentences we write down express the propositions that are in the argument. This is paraphrasing. Other times, we have to do even more: occasionally, we have to fill in missing propositions; argumentative passages might not state all of the propositions in an argument explicitly, and in the course of explicating their arguments, we have to make these implicit, tacit propositions explicit by writing down the appropriate declarative sentences.

There's a fancy Greek word for argumentative passages that leave certain propositions unstated: *enthymemes*. Here's an example:

Cicero has more experience in public office than Don Quixote; He has a much deeper knowledge of the issues; He's the only one with the proper temperament to lead our country. I rest my case.

Again, the argumentative intentions here are plain: this person is staking out a position on a controversial topic—a presidential election. But notice, that position—that one should prefer Cicero to Don Quixote — is never stated explicitly. We get *reasons* for having that preference—the premises of the argument are explicit—but we never get a statement of the conclusion. But since this is clearly the upshot of the passage, we need to include a sentence expressing it in our explication:

Cicero has more experience than Don Quixote.

Cicero has deeper knowledge of issues than Don Quixote.

Cicero has the proper temperament to lead the country, while Don Quixote does not.

/∴ One should prefer Cicero to Don Quixote in the presidential election.

In that example, the conclusion of the argument was tacit. Sometimes, premises are unstated and we should make them explicit in our explication of the argument. Now consider this passage:

The sad fact is that wages for middle-class workers have stagnated over the past several decades. We need a resurgence of the union movement in this country.

This person is arguing in favor of labor unions; the second sentence is the conclusion of the argument. The first sentence gives the only explicit premise: the stagnation of middle-class wages. But notice what the passage doesn't say: what connection there might be between the two things. What do unions have to do with middle-class wages?

There's an implicit premise lurking in the background here—something that hasn't been said, but which needs to be true for the argument to go through. We need a claim that connects the premise to the conclusion—that bridges the gap between them. Something like this: A resurgence of unions would lead to wage growth for middle-class workers. The first sentence identifies a problem; the second sentence purports to give a solution to the problem. But it's only a solution if the tacit premise we've uncovered is true. If unions don't help raise middle-class wages, then the argument falls apart.

This is the mark of the kinds of tacit premises we want to uncover: if they're false, they undermine the argument. Often, premises like this are unstated for a reason: they're controversial claims on their own, requiring a lot of evidence to support them; so the arguer leaves them out, preferring not to get bogged down. When we draw them out, however, we can force a more robust dialectical exchange, focusing the argument on the heart of the matter. In this case, a discussion about the connection between unions and middle-class wages would be in order. There's a lot to be said on that topic.

An important aspect of understanding arguments, especially enthymetic ones is applying the principle of charity to our interpretation. It is possible a person may deliver an argument that seems to be lacking or premise, or a premise is misspoken and thus fails to execute its function in the argument. However, for our own sakes we should operate from and, more importantly, wrestle with the strongest version of an argument that can be plausibly interpreted from the

presentations we encounter. Applying the principle of charity when we interpret arguments makes us better logicians and better at benefitting from sharing arguments.

## *Arguments vs. Explanations*

One final item on the topic of “Recognizing and Explicating Arguments.” We’ve been focusing on explication; this is a remark about the recognition side. Some passages may superficially resemble arguments—they may, for example, contain words like ‘therefore’ and ‘because’, which normally indicate conclusions and premises in argumentative passages—but which are nevertheless not argumentative. Instead, they are explanations. Consider this passage:

Because female authors of her time were often stereotyped as writing light-hearted romances, and because her real name was well-known for other (sometimes scandalous) reasons, Mary Ann Evans was reluctant to use her own name for her novels. She wanted her work to be taken seriously and judged on its own merits. Therefore, she adopted the pen name ‘George Eliot’.

This passage has the words ‘because’ (twice), and ‘therefore’, which typically indicate the presence of premises and a conclusion, respectively. But it is not an argument. It’s not an argument because it does not have the *rhetorical purpose* of an argument: the aim of the passage is not to convince you of something. If it were an argument, the conclusion would be the claim following ‘therefore’, namely, the proposition that Mary Ann Evans adopted the pen name ‘George Eliot’. But this claim is not the conclusion of an argument; the passage is not trying to persuade us to believe that Evans adopted a pen name. That she did so is not a controversial claim. Rather, that’s a fact that’s assumed to be known already. The aim of the passage is to *explain* to us *why* Evans made that choice. The rhetorical purpose is not to convince; it is to inform, to edify. The speaker wants to make a point about Evans and her circumstances clear to the reader. The passage is an explanation, not an argument.

So, to determine whether a given passage is an argument or an explanation, we need to figure out its rhetorical purpose. Why is the author saying these things to me? Is she trying to convince me of something, or is she merely trying to inform me—to give me an explanation for something I already knew? Sometimes this is easy, as with the George Eliot passage; it’s hard to imagine someone saying those things with persuasive intent. Other times, however, it’s not so easy. Consider the following:

Many of the celebratory rituals [of Christmas], as well as the timing of the holiday, have their origins outside of, and may predate, the Christian commemoration of the birth of Jesus. Those traditions, at their best, have much to do with celebrating human relationships and the enjoyment that this life has to offer. As an atheist, I have no hesitation in embracing the holiday and joining with believers and nonbelievers alike to celebrate what we have in common.<sup>4</sup>

Unless we understand a little bit more about the context of this passage, it’s difficult to determine the speaker’s intentions. It may appear to be an argument. That atheists should embrace a religious holiday like Christmas is, among many, a controversial claim. Controversial claims are the kinds of claims that we often try to convince skeptical people to believe. If the speaker’s audience for this passage is a bunch of hard-line atheists, who vehemently reject anything with a whiff of religiosity, who consider Christmas a humbug, then it’s pretty clear that the speaker is trying to offer reasons for them to reconsider their stance; he’s trying to convince them to embrace Christmas; he’s making an argument. If we explicated the argument, we would paraphrase the last sentence to represent the controversial conclusion: ‘Atheists should have no hesitation embracing and celebrating Christmas’.

But in a different context, with a different audience, this may not be an argument. If we leave the claim in the final sentence as-is—‘As an atheist, I have no hesitation in embracing the holiday and joining with believers and nonbelievers

4. John Teehan, 12/24/2006, “A Holiday Season for Atheists, Too,” The New York Times. Excerpted in Copi and Cohen, 2009, Introduction to Logic 13e, p. 25.

alike to celebrate what we have in common—we have a claim about the speaker’s personal beliefs and inclinations. Typically, as we saw above, such claims are not suitable as the conclusions of arguments; we don’t usually spend time trying to convince people *that* we believe such-and-such. But what is more typical is providing people with *explanations* for *why* we believe things. If the author of our passage is an atheist, and he’s saying these things to friends of his, say, who know he’s an atheist, we might have just such an explanation. His friends know he’s not religious, but they know he loves Christmas. That’s kind of weird. Don’t atheists hate religious holidays? Not so, says our speaker. Let me *explain* to you why I have no problems with Christmas, despite my atheism.

Again, the difference between arguments and explanations comes down to rhetorical purpose: arguments try to convince people; explanations try to inform them. Determining whether a given passage is one or the other involves figuring out the author’s intentions. To do this, we must carefully consider the context of the passage.

## EXERCISES

1. Identify the conclusions in the following arguments.

- (a) Every citizen has a right—nay, a duty—to defend himself and his family. This is all the more important in these increasingly dangerous times. The framers of the Constitution, in their wisdom, enshrined the right to bear arms in that very document. We should all oppose efforts to restrict access to guns.
- (b) Totino’s pizza rolls are the perfect food. They have all the great flavor of pizza, with the added benefit of portability!
- (c) Because they go overboard making things user-friendly, Apple phones are inferior to those with Android operating systems. If you want to change the default settings on an Apple phone to customize it to your personal preferences, it’s practically impossible to figure out how. The interface is so dumbed down to appeal to the “average consumer” that it’s super hard to find where the controls for advanced settings even are. On Android phones, though, everything’s right there in the open.
- (d) The U.S. incarcerates more people per capita than any other country on Earth, many for non-violent drug offenses. Militarized policing of our inner cities has led to scores of unnecessary deaths and a breakdown of trust between law enforcement and the communities they are supposed to serve and protect. We need to end the “War on Drugs” now. Our criminal justice system is broken. The War on Drugs broke it.
- (e) The point of a watch is to tell you what time it is. Period. Rolexes are a complete waste of money. They don’t do any better at telling the time, and they cost a ton!

2. Explicate the following arguments, paraphrasing as necessary.

- (a) You think destroying the ring of power is a good idea?! Why destroy the most powerful tool we have. It has come into our hands to aid us in our time of need. Without it, we will be too weak to defeat Sauron. I say we must not destroy, but rather, we must wield it.
- (b) The heat will escape the house through the open door, which means the heater will keep running, which will make our power bill go through the roof. Then we’ll be broke. So stop leaving the door open when you come into the house.
- (c) Do you like delicious food? How about fun games? And I know you like cool prizes. Well then, Chuck E. Cheese’s is the place for you.

3. Write down the tacit premises that the following arguments depend on for their success.

- (a) Cockfighting is an exciting pastime enjoyed by many people. It should therefore be legal.
- (b) The president doesn’t understand the threat we face. He won’t even use the phrase “Hyper-temporal Space Vampires.”

4. Write down the tacit conclusion that follows most immediately from the following.

- (a) If there really were an all-loving God looking down on us, then there wouldn’t be so much death and destruction visited upon innocent people.

(b) The death penalty is immoral. Numerous studies have shown that there is racial bias in its application. The rise of DNA testing has exonerated scores of inmates on death row; who knows how many innocent people have been killed in the past? The death penalty is also impractical. Revenge is counterproductive: “An eye for an eye leaves the whole world blind,” as Gandhi said. Moreover, the costs of litigating death penalty cases, with their endless appeals, are enormous. The correct decision for policymakers is clear.

5. Decide whether the following are arguments or explanations, given their context. If the passage is an argument, write down its conclusion; if it is an explanation, write down the fact that is being explained.

(a) Michael Jordan is the best of all time. I don’t care if Kareem scored more points; I don’t care if Russell won more championships. The simple fact is that no other player in history displayed the stunning combination of athleticism, competitive drive, work ethic, and sheer jaw-dropping artistry of Michael Jordan. [Context: Sports talk radio host going on a “rant”]

(b) Because different wavelengths of light travel at different velocities when they pass through water droplets, they are refracted at different angles. Because these different wavelengths correspond to different colors, we see the colors separated. Therefore, if the conditions are right, rainbows appear when the sun shines through the rain. [Context: grade school science textbook]

(d) This is how natural selection works: those species with traits that promote reproduction tend to have an advantage over competitors and survive; those without such traits tend to die off. The way that humans reproduce is by having sex. Since the human species has survived, it must have traits that encourage reproduction—that encourage having sex. This is why sex feels good. Sex feels good because if it didn’t, the species would not have survived. [Context: excerpt from *Evolutionary Biology for Dummies*]

## IV. Deductive and Inductive Arguments

As we noted earlier, there are different logics—different approaches to distinguishing good arguments from bad ones. One of the reasons we need different logics is that there are different kinds of arguments. In this section, we distinguish two types: deductive and inductive arguments.

### *Deductive Arguments*

First, deductive arguments. These are distinguished by their aim: a *deductive argument* attempts to provide premises that guarantee, necessitate its conclusion. Success for a deductive argument, then, does not come in degrees: either the premises do in fact guarantee the conclusion, in which case the argument is a good, successful one, or they don’t, in which case it fails. Evaluation of deductive arguments is a black-and-white, yes-or-no affair; there is no middle ground. We have a special term for a successful deductive argument: we call it *valid*. Validity is a central concept in the study of logic. It’s so important, we’re going to define it three times. Each of these three definitions is equivalent to the others; they are just three different ways of saying the same thing:

An argument is *valid* just in case...

- (i) its premises guarantee its conclusion; i.e.,
- (ii) IF its premises are true, then its conclusion must also be true; i.e.,
- (iii) it is impossible for its premises to be true and its conclusion false.

Here’s an example of a valid deductive argument:

All humans are mortal.

Socrates is a human.

/∴ Socrates is mortal.

Here is a second example:



If we go to a Chinese restaurant I'm having Kung Pao Tofu.  
If we go to a Mexican restaurant I'm have Pollo con Mole.  
We are either going to a Chinese restaurant of a Mexican restuarant.  
/  $\therefore$  I'm either having Kung Pao Tofu or Pollo con Mole.

These arguments are valid because the premises do in fact guarantee the conclusion: if they're true (as a matter of fact, they are), then the conclusions must be true; it's impossible for the premises to be true and the conclusions false. And we are only speaking of *if* premises were true *then* conclusion would be. We are not talking about whether the premises match the world, or not.

Here's a surprising fact about validity: what makes a deductive argument valid has nothing to do with its content; rather, validity is determined by the argument's *form*. That is to say, what makes our Socrates argument valid is not that it says a bunch of accurate things about Socrates, humanity, and mortality. The content doesn't make a difference. Instead, it's the form that matters—the pattern that the argument exhibits.

Later, when undertake a more detailed study of deductive logic, we will give a precise definition of logical form.<sup>5</sup> For now, we'll use this rough gloss: the form of an argument is what's left over when you strip away all the non-logical terms and replace them with blanks.

Here's what that looks like for our Socrates argument:

All A are B.  
x is A.  
/  $\therefore$  x is B.

The letter are the blanks: they're placeholders, variables. As a matter of convention, we're using capital letters to stand for groups of things (humans, mortals) and lower case letters to stand for individual things (Socrates).

The Socrates argument is a good, valid argument because it exhibits this good, valid form. Our third way of wording the definition of validity helps us see why this is a valid form: it's impossible for the premises to be true and the conclusion false, in that it's impossible to plug in terms for A, B, and x in such a way that the premises come out true and the conclusion comes out false.

A consequence of the fact that validity is determined entirely by an argument's form is that, given a valid form, every single argument that has that form will be valid. So any argument that has the same form as our Socrates argument will be valid; that is, we can pick things at random to stick in for A, B, and x, and we're guaranteed to get a valid argument. Here's a silly example:

All apples are bananas.  
Moby Dick is an apple.  
/  $\therefore$  Moby Dick is a banana.

This argument has the same form as the Socrates argument: we simply replaced A with 'apples', B with 'bananas', and x with 'Moby Dick'. That means it's a valid argument. That's a strange thing to say, since the argument is just silly—but it's the form that matters, not the content. Our second way of wording the definition of validity can help us here. The standard for validity is this: IF the premises are true, then the conclusion must be. That's a big 'IF'. In this case, as a matter of fact, the premises are not true (they're silly, plainly false). However, IF they were true—if in fact apples were a type of banana and Moby Dick were an apple—then the conclusion would be unavoidable: Moby would have to be a banana. The premises aren't true, but if they were, the conclusion would have to be—that's validity.

So it turns out that the actual truth or falsehood of the propositions in a valid argument are completely irrelevant to its validity. The Socrates argument has all true propositions and it's valid; the Moby Dick argument has all false propositions, but it's valid, too. They're both valid because they have a valid form; the truth/falsity of their propositions don't make any difference. This means that a valid argument can have propositions with almost any combination of

5. Definitions, actually. We'll study two different deductive logics, each with its own definition of form.

truth values: some true premises, some false ones, a true or false conclusion. One can fiddle around with the Socrates' argument's form, plugging different things in for A, B, and x, and see that this is so. For example, plug in 'ants' for A, 'bugs' for B, and Beyoncé for x: you get one true premise (All ants are bugs), one false one (Beyoncé is an ant), and a false conclusion (Beyoncé is a bug). Plug in other things and you can get any other combination of truth-values.

Any combination, that is, but one: you'll never get true premises and a false conclusion. That's because the Socrates' argument's form is a valid one; by definition, it's impossible to generate true premises and a false conclusion in that case.

This irrelevance of truth-value to judgments about validity means that those judgments are immune to revision. That is, once we decide whether an argument is valid or not, that decision cannot be changed by the discovery of new information. New information might change our judgment about whether a particular proposition in our argument is true or false, but that can't change our judgment about validity. Validity is determined by the argument's form, and new information can't change the form of an argument. The Socrates argument is valid because it has a valid form. Suppose we discovered, say, that as a matter of fact Socrates wasn't a human being at all, but rather an alien from outer space who got a kick out of harassing random people on the streets of ancient Athens. That information would change the argument's second premise—Socrates is human—from a truth to a falsehood. But it wouldn't make the argument invalid. The form is still the same, and it's a valid one.

It's time to face up to an awkward consequence of our definition of validity. Remember, logic is about evaluating arguments—saying whether they're good or bad. We've said that for deductive arguments, the standard for goodness is validity: the good deductive arguments are the valid ones. Here's where the awkwardness comes in: because validity is determined by form, it's possible to generate valid arguments that are nevertheless completely ridiculous-sounding on their face. Remember, the Moby Dick argument—where we concluded that he's a banana—is valid. In other words, we're saying that the Trump argument is good; it's valid, so it gets the logical thumbsup. But that's nuts! The Moby Dick argument is obviously bad, in some sense of 'bad', right? It's a collection of silly, nonsensical claims.

We need a new concept to specify what's wrong with the Moby Dick argument. That concept is *soundness*. This is a higher standard of argument-goodness than validity; in order to meet it, an argument must satisfy two conditions.

An argument is *sound* just in case (i) it's valid, AND (ii) its premises are in fact true.

The Moby Dick argument, while valid, is not sound, because it fails to satisfy the second condition: its premises are both false. The Socrates argument, however, which is valid and contains nothing but truths (Socrates was not in fact an alien), is sound. Since in all valid arguments the truth of the premises guarantee the truth of the conclusion, and because all sound arguments are also valid, in sound arguments the conclusions are true.

The question now naturally arises: if soundness is a higher standard of argument-goodness than validity, why didn't we say that in the first place? Why so much emphasis on validity? The answer is this: we're doing logic here, and as logicians, we have no special insight into the soundness of arguments. Or rather, we should say that as logicians, we have only partial expertise on the question of soundness. Logic can tell us whether or not an argument is valid, but it cannot tell us whether or not it is sound. Logic has no special insight into the second condition for soundness, the actual truth-values of premises. To take an example from the silly Moby Dick argument, suppose you weren't sure about the truth of the first premise, which claims that all apples are bananas (you have very little experience with fruit, apparently). How would you go about determining whether that claim was true or false? Whom would you ask? Well, this is a pretty easy one, so you could ask pretty much anybody, but the point is this: if you weren't sure about the relationship between apples and bananas, you wouldn't think to yourself, "I better go find a logician to help me figure this out." Propositions make claims about how things are in the world. To figure out whether they're true or false, you need to consult experts in the relevant subject-matter. Most claims aren't about logic, so logic is very little help in determining truth-values. Since logic can only provide insight into the validity half of the soundness question, we focus on validity and leave soundness to one side.

Returning to validity, then, we're now in a position to do some actual logic. Given what we know, we can *demonstrate invalidity*; that is, we can prove that an invalid argument is invalid, and therefore bad (it can't be sound, either; the first condition for soundness is validity, so if the argument's invalid, the question of actual truth-values doesn't even come up). Here's how:

To *demonstrate the invalidity* of an argument, one must write a down a new argument with the same form as the original, whose premises are in fact true and whose conclusion is in fact false. This new argument is called a *counterexample*. When we use it, we are giving a *refutation by logical analogy* because the new argument with same form is logically analogous to the argument under attack.

Let's look at an example. The following argument is invalid:

Some mammals are swimmers.  
All whales are swimmers.  
∴ All whales are mammals.

Now, it's not really obvious that the argument is invalid. It does have one thing going for it: all the claims it makes are true. But we know that doesn't make any difference, since validity is determined by the argument's form, not its content. If this argument is invalid, it's invalid because it has a bad, invalid form. This is the form:

Some A are B.  
All C are B.  
∴ All C are A.

To prove that the original whale argument is invalid, we have to show that this form is invalid. For a valid form, we learned, it's impossible to plug things into the blanks and get true premises and a false conclusion; so for an invalid form, it's *possible* to plug things into the blanks and get that result. That's how we generate our counterexample: we plug things in for A, B, and C so that the premises turn out true and the conclusion turns out false. There's no real method here; you just use your imagination to come up with an A, B, and C that give the desired result. Here's a counterexample:

Some lawyers are American citizens.  
All members of Congress are American citizens.  
∴ All members of Congress are lawyers.

For A, we inserted 'lawyers', for B we chose 'American citizens', and for C, 'members of Congress'. The first premise is clearly true. The second premise is true: non-citizens aren't eligible to be in Congress. And the conclusion is false: there are lots of people in Congress who are non-lawyers—doctors, businesspeople, etc.

That's all we need to do to prove that the original whale-argument is invalid: come up with one counterexample, one way of filling in the blanks in its form to get true premises and a false conclusion. We only have to prove that it's *possible* to get true premises and a false conclusion, and for that, you only need one example.

What's far more difficult is to prove that a particular argument is valid. To do that, we'd have to show that its form is such that it's *impossible* to generate a counterexample, to fill in the blanks to get true premises and a false conclusion. Proving that it's possible is easy; you only need one counterexample. Proving that it's impossible is hard; in fact, at first glance, it looks impossibly hard! What do you do? Check all the possible ways of plugging things into the blanks, and make sure that none of them turn out to have true premises and a false conclusion? That's nuts! There are, literally, infinitely many ways to fill in the blanks in an argument's form. Nobody has the time to check infinitely many potential counterexamples.

Well, take heart; it's still early. For now, we're able to do a little bit of deductive logic: given an invalid argument, we can demonstrate that it is in fact invalid. We're not yet in the position we'd like to be in, namely of being able to determine, for any argument whatsoever, whether it's valid or not. Proving validity looks too hard based on what we know so far. But we'll know more later: in chapters 3 and 4 we will study two deductive logics, and each one will give us a method of deciding whether or not any given argument is valid. But that'll have to wait. Baby steps.

## *Inductive Arguments*

That's all we'll say for now about deductive arguments. On to the other type of argument we're introducing in this section: *inductive arguments*. These are distinguished from their deductive cousins by their relative lack of ambition. Whereas deductive arguments aim to give premises that guarantee/necessitate the conclusion, inductive arguments are more modest: they aim merely to provide premises that make the conclusion more probable than it otherwise would

be; they aim to support the conclusion, but without making it unavoidable. Inductive arguments do not claiming the conclusion follows with absolute certainty.

Here is an example of an inductive argument:

I'm telling you, you're not going die taking a plane to visit us. Airplane crashes happen far less frequently than car crashes, for example; so you're taking a bigger risk if you drive. In fact, plane crashes are so rare, you're far more likely to die from slipping in the bathtub. You're not going to stop taking showers, are you?

The speaker is trying to convince her visitor that he won't die in a plane crash on the way to visit her. That's the conclusion: you won't die. This claim is supported by the others—which emphasize how rare plane crashes are—but it is not guaranteed by them. After all, plane crashes sometimes do happen. Instead, the premises give reasons to believe that the conclusion—you won't die—is very probable.

Since inductive arguments have a different, more modest goal than their deductive cousins, it would be unreasonable for us to apply the same evaluative standards to both kinds of argument. That is, we can't use the terms 'valid' and 'invalid' to apply to inductive arguments. Remember, for an argument to be valid, its premises must guarantee its conclusion. But inductive arguments don't even try to provide a guarantee of the conclusion; technically, then, they're all invalid. But that won't do. We need a different evaluative vocabulary to apply to inductive arguments. We will say of inductive arguments that they are (relatively) *strong* or *weak*, depending on how probable their conclusions are in light of their premises. One inductive argument is stronger than another when its conclusion is more probable than the other, given their respective premises.

One consequence of this difference in evaluative standards for inductive and deductive arguments is that for the former, unlike the latter, our evaluations are subject to revision in light of new evidence. Recall that since the validity or invalidity of a deductive argument is determined entirely by its form, as opposed to its content, the discovery of new information could not affect our evaluation of those arguments. The Socrates argument remained valid, even if we discovered that Socrates was in fact an alien. Our evaluations of inductive arguments, though, are not immune to revision in this way. New information might make the conclusion of an inductive argument more or less probable, and so we would have to revise our judgment accordingly, saying that the argument is stronger or weaker. Returning to the example above about plane crashes, suppose we were to discover that the FBI in the visitor's hometown had recently been hearing lots of "chatter" from terrorist groups active in the area, with strong indications that they were planning to blow up a passenger plane. Yikes! This would affect our estimation of the probability of the conclusion of the argument—that the visitor wasn't going to die in a crash. The probability of not dying goes down (as the probability of dying goes up). This new information would trigger a re-evaluation of the argument, and we would say it's now weaker. If, on the other hand, we were to learn that the airline that flies between the visitor's and the speaker's towns had recently upgraded its entire fleet, getting rid of all of its older planes, replacing them with newer, more reliable model, while in addition instituting a new, more thorough and rigorous program of pre- and post-flight safety and maintenance inspections—well, then we might revise our judgment in the other direction. Given this information, we might judge that things are even safer for the visitor as it regards plane travel; that is, the proposition that the visitor won't die is now even more probable than it was before. This new information would strengthen the argument to that conclusion.

Reasonable follow-up question: *how much* is the argument strengthened or weakened by the new information imagined in these scenarios? Answer: how should I know? Sorry, that's not very helpful. But here's the point: we're talking about probabilities here; sometimes it's hard to know what the probability of something happening really is. Sometimes it's not: if I flip a coin, I know that the probability of it coming up tails is 0.5. But how probable is it that a particular plane from Airline X will crash with our hypothetical visitor on board? I don't know. And how much more probable is a disaster on the assumption of increased terrorist chatter? Again, I have no idea. All I know is that the probability of dying on the plane goes up in that case. And in the scenario in which Airline X has lots of new planes and security measures, the probability of a crash goes down. Sometimes, with inductive arguments, all we can do is make *relative* judgments about strength and weakness: in light of these new facts, the conclusion is more or less probable than it was before we learned of the new facts. Sometimes, however, we can be precise about probabilities and make *absolute* judgments about strength and weakness: we can say precisely how probable a conclusion is in light of the premises supporting it.

But this is a more advanced topic. We will discuss inductive logic in chapters 5 and 6, and will go into more depth then. Until then, patience. Baby steps.

If our inductive arguments are based upon reasonably true and accurate premises and the reasoning is strong, then the conclusion is likely to be accurate. In this case, we would call the argument not only strong, but also *cogent*. Cogency is like the inductive analogue of soundness. While formal logic does not deal with soundness, informal logic and inductive reasoning does give us many considerations with which to weigh the cogency of claims and possibly suspect errors made in affirming them.

## EXERCISES

1. Determine whether the following statements are true or false.

- (a) Not all valid arguments are sound.
- (b) An argument with a false conclusion cannot be sound.
- (c) An argument with true premises and a true conclusion is valid.
- (d) An argument with a false conclusion cannot be valid.

2. Demonstrate that the following arguments are invalid.

(a)

Some politicians are soundrels.

Crazy Hank is a politician.

/∴ Crazy Hank is a soundrel.

The argument's form is:

Some A are B.

x is A.

/∴ x is B.

[where 'A' and 'B' stand for groups of things and 'x' stands for an individual]

(b)

All dinosaurs are animals.

Some animals are extinct.

/∴ All dinosaurs are extinct.

The argument's form is:

All A are B.

Some B are C.

/∴ All A are C.

[where 'A', 'B', and 'C' stand for groups of things]

3. Consider the following inductive argument (about a made-up person):

Sally Johansson does all her grocery shopping at an organic food co-op. She's a huge fan of tofu. She's really into those week-long juice cleanse thingies. And she's an active member of PETA. I conclude that she's a vegetarian.

- (a) Make up a new piece of information about Sally that weakens the argument.
- (b) Make up a new piece of information about Sally that strengthens the argument.

## V. Diagramming Arguments

Before we get down to the business of *evaluating* arguments—of judging them valid or invalid, strong or weak—we still need to do some preliminary work. We need to develop our analytical skills to gain a deeper understanding of how arguments are constructed, how they hang together. So far, we've said that the premises are there to support the conclusion. But we've done very little in the way of analyzing the structure of arguments: we've just separated the premises from the conclusion. We know that the premises are supposed to support the conclusion. What we haven't explored is the question of just *how* the premises in a given argument do that job—how they work together to support the conclusion, what kinds of relationships they have with one another. This is a deeper level of analysis than merely distinguishing the premises from the conclusion; it will require a mode of presentation more elaborate than a list of propositions with the bottom one separated from the others by a horizontal line. To display our understanding of the relationships among premises supporting the conclusion, we are going to *depict* them: we are going to draw diagrams of arguments.

Here's how the diagrams will work. They will consist of three elements: (1) circles with numbers inside them—each of the propositions in the argument we're diagramming will be assigned a number, so these circled numbers in the diagram will represent the propositions; (2) arrows pointed at circled numbers—these will represent relationships of support, where one or more propositions provide a reason for believing the one pointed to; and (3) horizontal brackets—propositions connected by these will be interdependent (in a sense to be specified below).

Our diagrams will always feature the circled number corresponding to the conclusion at the bottom. The premises will be above, with brackets and arrows indicating how they collectively support the conclusion and how they're related to one another. There are a number of different relationships that premises can have to one another. We will learn how to draw diagrams of arguments by considering them in turn.

### *Independent Premises*

Often, different premises will support a conclusion—or another premise—individually, without help from any others. When this is the case, we draw an arrow from the circled number representing that premise to the circled number representing the proposition it supports.

Consider this simple argument:

① Marijuana is less addictive than alcohol. In addition, ② it can be used as a medicine to treat a variety of conditions. Therefore, ③ marijuana should be legal.

The last proposition is clearly the conclusion (the word 'therefore' is a big clue), and the first two propositions are the premises supporting it. They support the conclusion *independently*. The mark of independence is this: each of the premises would still provide support for the conclusion even if the other weren't true; each, on its own, gives you a reason for believing the conclusion. In this case, then, we diagram the argument as follows:



### *Intermediate Premises*

Some premises support their conclusions more directly than others. Premises provide more indirect support for a conclusion by providing a reason to believe another premise that supports the conclusion more directly. That is, some premises are intermediate between the conclusion and other premises.

Consider this simple argument:

① Automatic weapons should be illegal. ② They can be used to kill large numbers of people in a short amount of time. This is because ③ all you have to do is hold down the trigger and bullets come flying out in rapid succession.

The conclusion of this argument is the first proposition, so the premises are propositions 2 and 3. Notice, though, that there's a relationship between those two claims. The third sentence starts with the phrase 'This is because', indicating that it provides a reason for another claim. The other claim is proposition 2; 'This' refers to the claim that automatic weapons can kill large numbers of people quickly. Why should I believe that they can do that? Because all one has to do is hold down the trigger to release lots of bullets really fast. Proposition 2 provides immediate support for the conclusion (automatic weapons can kill lots of people really quickly, so we should make them illegal); proposition 3 supports the conclusion more indirectly, by giving support to proposition 2. Here is how we diagram in this case:



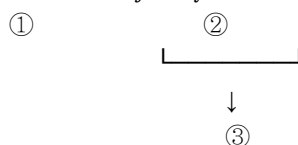
### *Joint Premises*

Sometimes premises need each other: the job of supporting another proposition can't be done by each on its own; they can only provide support together, jointly. Far from being independent, such premises are interdependent. In this situation, on our diagrams, we join together the interdependent premises with a bracket underneath their circled numbers.

There are a number of different ways in which premises can provide joint support. Sometimes, premises just fit together like a hand in a glove; or, switching metaphors, one premise is like the key that fits into the other to unlock the proposition they jointly support. An example can make this clear:

① The chef has decided that either salmon or chicken will be tonight's special. ② Salmon won't be the special. Therefore, ③ the special will be chicken.

Neither premise 1 nor premise 2 can support the conclusion on its own. A useful rule of thumb for checking whether one proposition can support another is this: read the first proposition, then say the word 'therefore', then read the second proposition; if it doesn't make any sense, then you can't draw an arrow from the one to the other. Let's try it here: "The chef has decided that either salmon or chicken will be tonight's special; therefore, the special will be chicken." That doesn't make any sense. What happened to salmon? Proposition 1 can't support the conclusion on its own. Neither can the second: "Salmon won't be the special; therefore, the special will be chicken." Again, that makes no sense. Why chicken? What about steak, or lobster? The second proposition can't support the conclusion on its own, either; it needs help from the first proposition, which tells us that if it's not salmon, it's chicken. Propositions 1 and 2 need each other; they support the conclusion jointly. This is how we diagram the argument:



The same diagram would depict the following argument:

① John Le Carre gives us realistic, three-dimensional characters and complex, interesting plots. ② Ian Fleming, on the other hand, presents an unrealistically glamorous picture of international espionage, and his plotting isn't what you'd call immersive. ③ Le Carre is a better author of spy novels than Fleming.

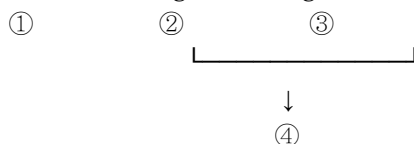
In this example, the premises work jointly in a different way than in the previous example. Rather than fitting together hand-in-glove, these premises each give us half of what we need to arrive at the conclusion. The conclusion is a

comparison between two authors. Each of the premises makes claims about one of the two authors. Neither one, on its own, can support the comparison, because the comparison is a claim about both of them. The premises can only support the conclusion together. We would diagram this argument the same way as the last one.

Another common pattern for joint premises is when *general* propositions need help to provide support for *particular* propositions. Consider the following argument:

- ① People shouldn't vote for racist, incompetent candidates for president. ② Loudmouth Larry seems to make a new racist remark at least twice a week. And ③ he lacks the competence to run even his own (failed) businesses, let alone the whole country. ④ You shouldn't vote for Larry to be the president.

The conclusion of the argument, the thing it's trying to convince us of, is the last proposition— you shouldn't vote for Loudmouth Larry. This is a particular claim: it's a claim about an individual person, Loudmouth Larry. The first proposition in the argument, on the other hand, is a general claim: it asserts that, generally speaking, people shouldn't vote for incompetent racists; it makes no mention of an individual candidate. It cannot, therefore, support the particular conclusion—about Loudmouth Larry—on its own. It needs help from other particular claims—propositions 2 and 3—that tell us that the individual in the conclusion, Loudmouth Larry, meets the conditions laid out in the general proposition 1: racism and incompetence. This is how we diagram the argument:



Occasionally, an argumentative passage will only explicitly state one of a set of joint premises because the others “go without saying”—they are part of the body of background information about which both speaker and audience agree. In the last example, that Loudmouth Larry was an incompetent racist was not uncontroversial background information. But consider this argument:

- ① It would be good for the country to have a woman with lots of experience in public office as president. ② People should vote for Cleopatra.

Diagramming this argument seems straightforward: an arrow pointing from ① to ②. But we've got the same relationship between the premise and conclusion as in the last example: the premise is a general claim, mentioning no individual at all, while the conclusion is a particular claim about Cleopatra. Doesn't the general premise “need help” from particular claims to the effect that the individual in question, Cleopatra, meets the conditions set forth in the premise—i.e., that she's a woman and that she has lots of experience in public office? No, not really. Everybody knows those things about her already; they go without saying, and can therefore be left unstated (implicit, tacit).

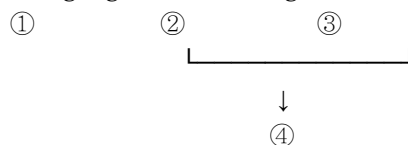
But suppose we had included those obvious truths about Cleopatra in our presentation of the argument; suppose we had made the tacit premises explicit:

- ① It would be good for the country to have a woman with lots of experience in public office as president.  
 ② Cleopatra is a woman. And ③ she has deep experience with public offices—as an empress, a diplomat, and military commander. ④ People should vote for Cleopatra.

How do we diagram this? Earlier, we talked about a rule of thumb for determining whether or not it's a good idea to draw an arrow from one number to another in a diagram: read the sentence corresponding to the first number, say the word ‘therefore’, then read the sentence corresponding to the second number; if it doesn't make sense, then the arrow is a bad idea. But if it does make sense, does that mean you *should* draw the arrow? Not necessarily. Consider the first and last sentences in this passage. Read the first, then ‘therefore’, then the last. Makes pretty good sense! That's just the original formulation of the argument with the tacit propositions remaining implicit. And in that case we said it would be OK to draw an arrow from the general premise's number straight to the conclusion's. But when we add the tacit premises—the second and third sentences in this passage—we can't draw an arrow directly from ① to ④. To do so would obscure the relationship among the first three propositions and misrepresent how the argument works. If we drew an arrow from ① to ④, what would we do with ② to ③ in our diagram? Do they get their own arrows, too? No, that won't do. Such a diagram would be telling us that the first three propositions each independently provide a reason for the conclusion. But they're clearly not independent; there's a relationship among them that our diagram must capture,



and it's the same relationship we saw in the parallel argument about Loudmouth Larry, with the particular claims in the second and third propositions working together with the general claim in the first:



The arguments we've looked at thus far have been quite short—only two or three premises. But of course some arguments are longer than that. Some are much longer. It may prove instructive, at this point, to tackle one of these longer bits of reasoning. It comes from the (fictional) master of analytical deductive reasoning, Sherlock Holmes. The following passage is from the first Holmes story—*A Study in Scarlet*, one of the few novels Arthur Conan Doyle wrote about his most famous character—and it's a bit of early dialogue that takes place shortly after Holmes and his longtime associate Dr. Watson meet for the first time. At that first meeting, Holmes did his typical Holmesy thing, where he takes a quick glance at a person and then immediately makes some startling inference about them, stating some fact about them that it seems impossible he could have known. Here they are—Holmes and Watson—talking about it a day or two later. Holmes is the first to speak:

“Observation with me is second nature. You appeared to be surprised when I told you, on our first meeting, that you had come from Afghanistan.”

“You were told, no doubt.”

“Nothing of the sort. I knew you came from Afghanistan. From long habit the train of thoughts ran so swiftly through my mind, that I arrived at the conclusion without being conscious of intermediate steps. There were such steps, however. The train of reasoning ran, ‘Here is a gentleman of a medical type, but with the air of a military man. Clearly an army doctor, then. He has just come from the tropics, for his face is dark, and that is not the natural tint of his skin, for his wrists are fair. He has undergone hardship and sickness, as his haggard face says clearly. His left arm has been injured. He holds it in a stiff and unnatural manner. Where in the tropics could an English army doctor have seen much hardship and got his arm wounded? Clearly in Afghanistan.’ The whole train of thought did not occupy a second. I then remarked that you came from Afghanistan, and you were astonished.”<sup>6</sup>

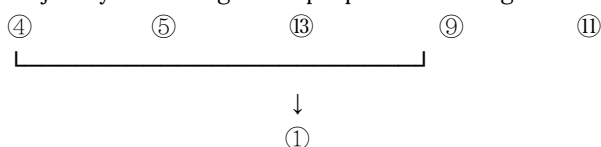
This is an extended inference, with lots of propositions leading to the conclusion that Watson had been in Afghanistan. Before we draw the diagram, let's number the propositions involved in the argument:

1. Watson was in Afghanistan.
2. Watson is a medical man.
3. Watson is a military man.
4. Watson is an army doctor.
5. Watson has just come from the tropics.
6. Watson's face is dark.
7. Watson's skin is not naturally dark.
8. Watson's wrists are fair.
9. Watson has undergone hardship and sickness.
10. Watson's face is haggard.
11. Watson's arm has been injured.
12. Watson holds his arm stiffly and unnaturally.
13. Only in Afghanistan could an English army doctor have been in the tropics, seen much hardship and got his arm wounded.

6. Excerpted in Copi and Cohen, 2009, *Introduction to Logic* 13e, pp. 58 - 59.

Lots of propositions, but they're mostly straightforward, right from the text. We just had to do a bit of paraphrasing on the last one—Holmes asks a rhetorical question and answers it, the upshot of which is the general proposition in 13. We know that proposition 1 is our conclusion, so that goes at the bottom of the diagram. The best thing to do is to start there and work our way up. Our next question is: Which premise or premises support that conclusion most directly? What goes on the next level up on our diagram?

It seems fairly clear that proposition 13 belongs on that level. The question is whether it is alone there, with an arrow from 13 to 1, or whether it needs some help. The answer is that it needs help. This is the general/particular pattern we identified above. The conclusion is about a particular individual—Watson. Proposition 13 is entirely general (presumably Holmes knows this because he reads the paper and knows the disposition of Her Majesty's troops throughout the Empire); it does not mention Watson. So proposition 13 needs help from other propositions that give us the relevant particulars about the individual, Watson. A number of conditions are laid out that a person must meet in order for us to conclude that they've been in Afghanistan: army doctor, being in the tropics, undergoing hardship, getting wounded. That Watson satisfies these conditions is asserted by, respectively, propositions 4, 5, 9, and 11. Those are the propositions that must work jointly with the general proposition 13 to give us our particular conclusion about Watson:



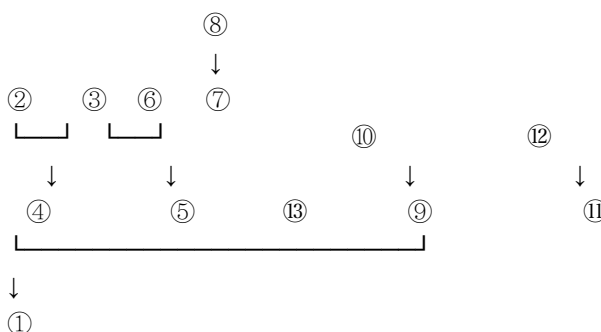
Next, we must figure out how what happens at the next level up. How are propositions 4, 5, 13, 9, and 11 justified? As we noted, the justification for 13 happens off-screen, as it were. Holmes is able to make that generalization because he follows the news and knows, presumably, that the only place in the British Empire where army troops are actively fighting in tropics is Afghanistan. The justification for the other propositions, however, is right there in the text.

Let's take them one at a time. First, proposition 4: Watson is an army doctor. How does Holmes support this claim? With propositions 2 and 3, which tell us that Watson is a medical and a military man, respectively. This is another pattern we've identified: these two proposition jointly support 4, because they each provide half of what we need to get there. There are two parts to the claim in 4: army and doctor. 2 gives us the doctor part; 3 gives us the army part. 2 and 3 jointly support 4.

Skipping 5 (it's a bit more involved), let's turn to 9 and 11, which are easily dispatched. What's the reason for believing 9, that Watson has suffered hardship? Go back to the passage. It's his haggard face that testifies to his suffering. Proposition 10 supports 9. Now 11: what evidence do we have that Watson's arm has been injured? Proposition 12: he holds it stiffly and unnaturally. 12 supports 11.

Finally, proposition 5: Watson was in the tropics. There are three propositions involved in supporting this one: 6, 7, and 8. Proposition 6 tells us Watson's face is dark; 7 tells us that his skin isn't naturally dark; 8 tells us his wrists are fair (light-colored skin). It's tempting to think that 6 on its own—dark skin—supports the claim that he was in the tropics. But it does not. One can have dark skin and not visited the tropics, provided one's skin is naturally dark. What tells us Watson has been in the tropics is that he has a tan—that his skin is dark *and* that's not its natural tone. 6 and 7 jointly support 5. And how do we know Watson's skin isn't naturally dark? By checking his wrists, which are fair: proposition 8 supports 7.

So this is our final diagram:



And there we go. An apparently unwieldy passage—thirteen propositions!—turns out not to be so bad. The lesson is that we must go step by step: start by identifying the conclusion, then ask which proposition(s) most directly support it; from there, work back until all the propositions have been diagrammed. Every long argument is just composed out of smaller, easily analyzed inferences.

## EXERCISES

Diagram the following arguments.

1. ① Gandalf would make a better president than Malfoy. ② Gandalf is a tough-minded pragmatist who gets things done. ③ Malfoy is a thin-skinned maniac who will be totally ineffective in dealing with Congress.
2. ① Numbskull is a jerk who's always offending people. Furthermore, ② he has no experience whatsoever in government. ③ Nobody should vote for Numbskull to be president.
3. ① Human beings evolved to eat meat, so ② eating meat is not immoral. ③ It's never immoral for a creature to act according to its evolutionary instincts.
4. ① We need new campaign finance laws in this country. ② The influence of Wall Street money on elections is causing a breakdown in our democracy with bad consequences for social justice. ③ Politicians who have taken those donations are effectively bought and paid for, consistently favoring policies that benefit the rich at the expense of the vast majority of citizens.
5. ① Voters shouldn't trust any politician who took money from Wall Street bankers. ② Drako accepted hundreds of thousands of dollars in speaking fee from Goldman Sachs, a big Wall Street firm. ③ You shouldn't trust him.
6. ① There are only three possible explanations for the presence of the gun at the crime scene: either the defendant just happened to hide from the police right next to where the gun was found, or the police planted the gun there after the fact, or it was really the defendant's gun like the prosecution says. ② The first option is too crazy a coincidence to be at all believable, and ③ we've been given no evidence at all that the officers on the scene had any means or motivation to plant the weapon. Therefore, ④ it has to be the defendant's gun.
7. ① Golden State has to be considered the clear favorite to win the NBA Championship. ② No team has ever lost in the Finals after taking a 3-games-to-1 lead, and ③ Golden State now leads Cleveland 3-to-1. In addition, ④ Golden State has the MVP of the league, Stephen Curry.
8. ① We should increase funding to public colleges and universities. First of all, ② as funding has decreased, students have had to shoulder a larger share of the financial burden of attending college, amassing huge amounts of debt. ③ A recent report shows that the average college student graduates with almost \$30,000 in debt. Second, ④ funding public universities is a good investment. ⑤ Every economist agrees that spending on public colleges is a good investment for states, where the economic benefits far outweigh the amount spent.
9. ① LED lightbulbs last for a really long time and ② they cost very little to keep lit. ③ They are, therefore, a great way to save money. ④ Old-fashioned incandescent bulbs, on the other hand, are wasteful. ⑤ You should buy LEDs instead of incandescent bulbs.
10. ① There's a hole in my left shoe, which means ② my feet will get wet when I wear them in the rain, and so ③ I'll probably catch a cold or something if I don't get a new pair of shoes. Furthermore, ④ having new shoes would make me look cool. ⑤ I should buy new shoes.
11. Look, it's just simple economics: ① if people stop buying a product, then companies will stop producing it. And ② people just aren't buying tablets as much anymore. ③ The CEO of Best Buy recently said that sales of tablets are "crashing" at his stores. ④ Samsung's sales of tablets were down 14% this year alone. ⑤ Apple's not going to continue to make your beloved iPad for much longer.
12. ① We should increase infrastructure spending as soon as possible. Why? First, ② the longer we delay needed repairs to things like roads and bridges, the more they will cost in the future. Second, ③ it would cause a

drop in unemployment, as workers would be hired to do the work. Third, ④ with interest rates at all-time lows, financing the spending would cost relatively little. A fourth reason? ⑤ Economic growth. ⑥ Most economists agree that government spending in the current climate would boost GDP.

13. ① Smoking causes cancer and ② cigarettes are really expensive. ③ You should quit smoking. ④ If you don't, you'll never get a girlfriend. ⑤ Smoking makes you less attractive to girls: ⑥ it stains your teeth and ⑦ it gives you bad breath.

14. ① The best cookbooks are comprehensive, well-written, and most importantly, have recipes that work. This is why ② Mark Bittman's classic *How to Cook Everything* is among the best cookbooks ever written. As its title indicates, ③ Bittman's book is comprehensive. Of course it doesn't literally teach you how to cook everything, but ④ it features recipes for cuisines from around the world—from French, Italian, and Spanish food to dishes from the Far and Middle East, as well as classic American comfort foods. In addition, ⑤ he covers almost every ingredient imaginable, with all different kinds of meats—including game—and every fruit and vegetable under the sun. ⑥ The book is also extremely well-written. ⑦ Bittman's prose is clear, concise, and even witty. Finally, ⑧ Bittman's recipes simply work. ⑨ In my many years of consulting *How to Cook Everything*, I've never had one lead me astray.

15. ① Logic teachers should make more money than CEOs. ② Logic is more important than business. ③ Without logic, we wouldn't be able to tell when people were trying to fool us: ④ we wouldn't know a good argument from a bad one. ⑤ But nobody would miss business if it went away. ⑥ What do businesses do except take our money? ⑦ And all those damned commercials they make; everybody hates commercials. ⑧ In a well-organized society, members of more important professions would be paid more, because ⑨ paying people is a great way to encourage them to do useful things. ⑩ People love money.

## 2.

MATTHEW KNACHEL

### Informal Logical Fallacies

#### I. Logical Fallacies: Formal and Informal

Generally and crudely speaking, a logical fallacy is just a bad argument. Bad, that is, in the logical sense of being incorrect—not bad in sense of being ineffective or unpersuasive. Alas, many fallacies are quite effective in persuading people; that is why they're so common. Often, they're not used mistakenly, but intentionally—to fool people, to get them to believe things that maybe they shouldn't. The goal of this chapter is to develop the ability to recognize these bad arguments for what they are so as not to be persuaded by them.

There are many, many fallacies out there. This chapter will examine the most common ones. By the end of this chapter, you should feel familiar with these villains. Throughout the chapter, ask yourself, “can the same line of reasoning be applied to different plausible premises but lead to a conclusion that is obviously false?” If so, there's likely a fallacy sabotaging the works. Using our sword of appealing to logical analogy can help us cut through confounding trickery to expose poor reasoning.

There are formal and informal logical fallacies. The formal fallacies are simple: they're just invalid deductive arguments. There is a clear line here. Either the premises necessitate the conclusion, or they do not. Consider the following:

If the Orcs take the village of Maedhros then the buildings will fall down.

But the Orcs won't take Maedhros.

$\therefore$  The buildings won't fall down.

This argument is invalid. It's got an invalid form: If A then B; not A; therefore, not B. Any argument of this form is fallacious, an instance of “Denying the Antecedent.”<sup>1</sup> Just because A doesn't happen doesn't mean B mightn't occur from other reasons. We can leave it as an exercise for the reader to fill in propositions for A and B to get true premises and a false conclusion. Intuitively, it's possible for that to happen: for example, maybe the buildings in Maedhros will fall down for some other reason besides orc attack. “Deny the Antecedent” is a fallacy that occurs due to a faulty form of the argument. It is, therefore, known as a “formal fallacy.” We will return to this class of error later.

Our concern in this chapter is not with formal fallacies—arguments that are bad because they have a bad form—but with informal fallacies. These arguments are bad, roughly, because of their content. More than that: their content, context, and/or mode of delivery.

Consider Hitler. Here's a guy who convinced a lot of people to believe things they had no business believing (because they were false). How did he do it? With lots of fallacious arguments. But it wasn't just the contents of the arguments (appeals to fear and patriotism, personal attacks on opponents, etc.) that made them fallacious; it was also the context in which he made them, and the (extremely effective) way he delivered them. Leni Riefenstahl's famous 1935 documentary/propaganda film *Triumph of the Will*, which follows Hitler during a Nazi party rally in Nuremberg, illustrates this. It has lots of footage of Hitler giving speeches. We hear the jingoistic slogans and vitriolic attacks—but we also see important elements of his persuasive technique. First, the setting. We see Hitler marching through row upon row of

1. If/then propositions like the first premise are called “conditional” propositions. The A part is the so-called “antecedent” of the conditional. The second premise denies it, and an invalid conclusion is drawn from this denial. We will present more analysis of arguments with this sort of technical vocabular in chapters 4 and 5.

neatly formed and impeccably outfitted German troops—thousands of them—approaching a massive raised dais, behind which are stories-high banners with the swastika on a red field. The setting, the context for Hitler’s speeches, was literally awesome—designed to inspire awe. It makes his audience all the more receptive to his message, all the more persuadable. Moreover, Hitler’s speechifying technique was masterful. He is said to have practiced assiduously in front of a mirror, and it shows. His array of hand gestures, facial contortions, and vocal modulations were all expertly designed to have maximum impact on the audience.

This consideration of Hitler highlights a couple of important things about the informal fallacies. First, they’re more than just bad arguments—they’re rhetorical tricks, extra-logical techniques used intentionally to try to convince people of things they maybe ought not to believe. Second, they work! Hitler convinced an entire nation to believe all sorts of crazy things. And advertisers and politicians continue to use these same techniques all the time. It’s incumbent upon a responsible citizen and consumer to be aware of this, and to do everything possible to avoid being bamboozled. That means learning about the fallacies. Hence, this chapter.

There are lots of different informal logical fallacies, lots of different ways of defining and characterizing them, lots of different ways of organizing them into groups. Since Aristotle first did it in his *Sophistical Refutations*, authors of logic books have been defining and classifying the informal fallacies in various ways. These remarks are offered as a kind of disclaimer: the reader is warned that the particular presentation of the fallacies in this chapter will be unique and will disagree in various ways with other presentations, reflecting as it must the author’s own idiosyncratic interests, understanding, and estimation of what is important. This is as it should be and always is. The interested reader is encouraged to consult alternative sources for further edification.

We will discuss 20 different informal fallacies, and we will group them into four families: (1) Fallacies of Distraction, (2) Fallacies of Weak Induction, (3) Fallacies of Illicit Presumption, and (4) Fallacies of Linguistic Emphasis. We take these up in turn.

## II. Fallacies of Distraction

We will discuss five informal fallacies under this heading. What they all have in common is that they involve arguing in such a way that issue that’s supposed to be under discussion is somehow sidestepped, avoided, or ignored. These fallacies are often called “Fallacies of Relevance” because they involve arguments that are bad insofar as the reasons given are irrelevant to the issue at hand. People who use these techniques with malicious intent are attempting to distract their audience from the central questions they’re supposed to be addressing, allowing them to appear to win an argument that they haven’t really engaged in.

### *Appeal to The People (Argumentum ad Populum)*

The Latin name of this fallacy literally means “argument to the people,” where ‘the people’ is used in the pejorative sense of “the unwashed masses,” or “the fickle mob”—the *hoi polloi*.<sup>2</sup> It’s notoriously effective to play on people’s emotions to get them to go along with you, and that’s the technique identified here. But, the thought is, we shouldn’t decide whether or not to believe things based on an emotional response; emotions are a distraction, blocking hard-headed, rational analysis.

Go back to Hitler for a minute. He was an expert at the appeal to emotion. He played on Germans’ fears and prejudices,

2. Many of the fallacies have Latin names, because, as we noted, identifying the fallacies has been an occupation of logicians since ancient times, and because ancient and medieval work comes down to us in Latin, which was the language of scholarship in the West for centuries.

their economic anxieties, their sense of patriotism and nationalistic pride. He stoked these emotions with explicit denunciations of Jews and non-Germans, promises of the return of glory for the Fatherland—but also using the sorts of techniques we canvassed above, with awesome settings and hyper-sensational speechifying.

There are as many different versions of the appeal to emotion as there are human emotions. Fear is perhaps the most commonly exploited emotion for politicians. Political ads inevitably try to suggest to voters that one's opponent will take away medical care or leave us vulnerable to terrorists, or some other scary outcome—usually without a whole lot in the way of substantive proof that these fears are at all reasonable. This is a fallacious appeal to emotion.

Advertisers do it, too. Think of all the ads with sexy models schilling for cars or beers or whatever. What does sexiness have to do with how good a beer tastes? Nothing. The ads are trying to engage your emotions to get you thinking positively about their product. To imagine how irrelevant sexy models are to beer quality, imagine how funny it'd be to see an ad firms try the same trick of using models to sell toilet paper!

An extremely common technique, especially for advertisers, is to appeal to people's underlying desire to fit in, to be hip to what everybody else is doing, not to miss out. This is the *bandwagon appeal*. The advertisement assures us that a certain television show is #1 in the ratings—with the tacit conclusion being that we should be watching, too. But this is a fallacy. We've all known it's a fallacy since we were little kids, the first time we did something wrong because all of our friends were doing it, too, and our moms asked us, "If all of your friends jumped off a bridge, would you do that too?"

One more example: suppose you're one of those sleazy personal injury lawyers—an "ambulance chaser". You've got a client who was grocery shopping at Wal-Mart, and in the produce aisle she slipped on a grape that had fallen on the floor and injured herself. Your eyes turn into dollar signs and a cha-ching noise goes off in your brain: Wal-Mart has deep pockets. So on the day of the trial, what do you do? How do you coach your client? Tell her to wear her nicest outfit, to look her best? Of course not! You wheel her into the courtroom in a wheelchair (whether she needs it or not); you put one of those foam neck braces on her, maybe give her an eye patch for good measure. You tell her to periodically emit moans of pain. When you're summing up your case before the jury, you spend most of your time talking about the horrible suffering your client has undergone since the incident in the produce aisle: the hospital stays, the grueling physical therapy, the addiction to pain medications, etc., etc.

All of this is a classic fallacious appeal to emotion—specifically, in this case, pity. The people you're trying to convince are the jurors. The conclusion you have to convince them of, presumably, is that Wal-Mart was negligent and hence legally liable in the matter of the grape on the floor. The details don't matter, but there are specific conditions that have to be met—proved beyond a reasonable doubt—in order for the jury to find Wal-Mart guilty. But you're not addressing those (probably because you can't). Instead, you're trying to distract the jury from the real issue by playing to their emotions. You're trying to get them feeling sorry for your client, in the hopes that those emotions will cause them to bring in the verdict you want. That's why the appeal to emotion is a Fallacy of Distraction: the goal is to divert your attention from the dispassionate evaluation of premises and the degree to which they support their conclusion, to get you thinking with your heart instead of your brain.

### *Appeal to Force (Argumentum ad Baculum)*

In Latin, 'baculus' refers to a stick or a club, which you could clobber someone with, presumably. Perhaps the least subtle of the fallacies is the appeal to force, in which you attempt to convince your interlocutor to believe something by threatening him. Threats pretty clearly distract one from the business of dispassionately appraising premises' support for conclusions, so it's natural to classify this technique as a Fallacy of Distraction.

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

The appeal to force is can also be subtle. There is a very common, very effective debating technique that belongs under this heading, one that is a bit less overt than explicitly threatening someone who fails to share your opinions. It involves

the sub-conscious, rather than conscious, perception of a threat. Here's what you do: during the course of a debate, make yourself physically imposing; sit up in your chair, move closer to your opponent, use hand gestures, like pointing right in their face; cut them off in the middle of a sentence, shout them down, be angry and combative. If you do these things, you're likely to make your opponent very uncomfortable—physically and emotionally. They might start sweating a bit; their heart may beat a little faster. They'll get flustered and maybe trip over their words. They may lose their train of thought; winning points they may have made in the debate will come out wrong or not at all. You'll look like the more effective debater, and the audience's perception will be that you made the better argument.

But you didn't. You came off better because your opponent was uncomfortable. The discomfort was not caused by an actual threat of violence; on a conscious level, they never believed you were going to attack them physically. But you behaved in a way that triggered, at the sub-conscious level, the types of physical/emotional reactions that occur in the presence of an actual physical threat. This is the more subtle version of the appeal to force. It's very effective and quite common, especially on some cable news channels.

### *Straw Man*

This fallacy involves the misrepresentation of an opponent's viewpoint—an exaggeration or distortion of it that renders it indefensible, something nobody in their right mind would agree with. You make your opponent out to be a complete wacko (even though he isn't), then declare that you don't agree with his (made-up) position. Thus, you merely *appear* to defeat your opponent: your real opponent doesn't hold the crazy view you imputed to him; instead, you've defeated a distorted version of him, one of your own making, one that is easily dispatched. Instead of taking on the real man, you construct one out of straw, thrash it, and pretend to have achieved victory. It works if your audience doesn't realize what you've done, if they believe that your opponent really holds the crazy view.

Politicians are most frequently victims (and practitioners) of this tactic. After his 2005 State of the Union Address, President George W. Bush's proposals were treated with straw-man abuse such as:

George W. Bush's State of the Union Address, masked in talk of "freedom" and "democracy," was an outline of a brutal agenda of endless war, global empire, and the destruction of what remains of basic social services.<sup>3</sup>

Well, who's not against "endless war" and "destruction of basic social services"? That Bush guy must be a complete nut! But of course this characterization is a gross exaggeration of what was actually said in the speech, in which Bush declared that we must "confront regimes that continue to harbor terrorists and pursue weapons of mass murder" and rolled out his proposal for privatization of Social Security accounts. Whatever you think of those actual policies, you need to do more to undermine them than to mis-characterize them as "endless war" and "destruction of social services." That's distracting your audience from the real substance of the issues.

In 2009, during the (interminable) debate over President Obama's healthcare reform bill—the Patient Protection and Affordable Care Act—former vice presidential candidate Sarah Palin took to Facebook to denounce the bill thus:

The America I know and love is not one in which my parents or my baby with Down Syndrome will have to stand in front of Obama's "death panel" so his bureaucrats can decide, based on a subjective judgment of their "level of productivity in society," whether they are worthy of health care. Such a system is downright evil.

Yikes! That sounds like the vilest bill in the history of evil! Bureaucrats euthanizing Down Syndrome babies and their grandparents? Holy Cow. 'Death panel' and 'level of productivity in society' are even in quotes. Did she pull those phrases from the text of the bill?

Of course she didn't. This is a completely insane distortion of what's actually in the bill (the kernel of truth behind the "death panels" thing seems to be a provision in the Act calling for Medicare to fund doctor-patient conversations about end-of-life care); the non-partisan fact-checking outfit Politifact named it their "Lie of the Year" in 2009. Palin

3. International Action Center, Feb. 4 2005, <http://iacenter.org/folder06/stateoftheunion.htm>



is not taking on the bill or the president themselves; she's confronting a made-up version, defeating it (which is easy, because the madeup bill is evil as heck; I can't get the disturbing idea of a Kafkaesque Death Panel out of my head), and pretending to have won the debate. But this distraction only works if her audience believes her straw man is the real thing. Alas, many did. But of course this is why these techniques are used so frequently: they work.

## *Red Herring*

This fallacy gets its name from the actual fish. When herring are smoked, they turn red and are quite pungent. Stinky things can be used to distract hunting dogs, who of course follow the trail of their quarry by scent; if you pass over that trail with a stinky fish and run off in a different direction, the hound may be distracted and follow the wrong trail. Whether or not this practice was ever used to train hunting dogs, as some suppose, the connection to logic and argumentation is clear. One commits the red herring fallacy when one attempts to distract one's audience from the main thread of an argument, taking things off in a different direction. The diversion is often subtle, with the detour starting on a topic closely related to the original—but gradually wandering off into unrelated territory. The tactic is often (but not always) intentional: one commits the red herring fallacy because one is not comfortable arguing about a particular topic on the merits, often because one's case is weak; so instead, the arguer changes the subject to an issue about which he feels more confident, makes strong points on the new topic, and pretends to have won the original argument.<sup>4</sup>

A fictional example can illustrate the technique. Consider Frank, who, after a hard day at work, heads to the tavern to unwind. He has far too much to drink, and, unwisely, decides to drive home. Well, he's swerving all over the road, and he gets pulled over by the police. Let's suppose that Frank has been pulled over in a posh suburb where there's not a lot of crime. When the police officer tells him he's going to be arrested for drunk driving, Frank becomes belligerent:

"Where do you get off? You're barely even real cops out here in the 'burbs. All you do is sit around all day and pull people over for speeding and stuff. Why don't you go investigate some real crimes? There's probably some unsolved murders in the inner city they could use some help with. Why do you have to bother a hard-working citizen like me who just wants to go home and go to bed?"

Frank is committing the red herring fallacy (and not very subtly). The issue at hand is whether or not he deserves to be arrested for driving drunk. He clearly does. Frank is not comfortable arguing against that position on the merits. So he changes the subject—to one about which he feels like he can score some debating points. He talks about the police out here in the suburbs, who, not having much serious crime to deal with, spend most of their time issuing traffic violations. Yes, maybe that's not as taxing a job as policing in the city. Sure, there are lots of serious crimes in other jurisdictions that go unsolved. But that's beside the point! It's a distraction from the real issue of whether Frank should get a DUI.

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

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

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

4. People often offer red herring arguments unintentionally, without the subtle deceptive motivation to change the subject—usually because they're just parroting a red herring argument they heard from someone else. Sometimes a person's response will be off-topic, apparently because they weren't listening to their interlocutor or they're confused for some reason. I prefer to label such responses as instances of Missing the Point (*Ignoratio Elenchi*), a fallacy that some books discuss at length, but which I've just relegated to a footnote.

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

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

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

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

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

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

### *Argumentum ad Hominem*

Everybody always used the Latin for this one—usually shortened to just '*ad hominem*', which means 'at the person'. You commit this fallacy when, instead of attacking your opponent's views, you attack your opponent himself.

This fallacy comes in a lot of different forms; there are a lot of different ways to attack a person while ignoring (or downplaying) their actual arguments. To organize things a bit, we'll divide the various *ad hominem* attacks into two groups: Abusive and Circumstantial.

Abusive *ad hominem* is the more straightforward of the two. The simplest version is simply calling your opponent names instead of debating him. If you pepper your descriptions of your opponent with tendentious, unflattering, politically charged language, you can get a rhetorical leg-up and undermine the credibility of the opponent.

Another abusive *ad hominem* attack is *guilt by association*. Here, you tarnish your opponent by associating him or his views with someone or something that your audience despises. Again, cable news is full of tenuous comparisons of contemporary leaders and ideas to unpopular foils such as Nazis and the Bolsheviks.<sup>5</sup> Not every comparison like this

5. Comparing your opponent to Hitler—or the Nazis—is quite common. Some clever folks came up with a fake-Latin term for the tactic: *Argumentum ad Nazium* (cf. the real Latin phrase, *ad nauseum*—to the point of nausea). Such comparisons are so common that author Mike Godwin formulated "Godwin's Law of Nazi Analogies: As an online

is fallacious, of course. But in many cases, where the connection is particularly flimsy, someone is clearly pulling a fast one.

The circumstantial *ad hominem* fallacy is not as blunt an instrument as its abusive counterpart. It also involves attacking one's opponent, focusing on some aspect of his person—his circumstances—as the core of the criticism. This version of the fallacy comes in many different forms, and some of the circumstantial criticisms involved raise legitimate concerns about the relationship between the arguer and his argument. They only rise (sink?) to the level of fallacy when these criticisms are taken to be definitive refutations, which, on their own, they cannot be.

To see what we're talking about, consider the circumstantial *ad hominem* attack that points out one's opponent's *self-interest* in making the argument he does. Consider:

A recent study from scientists at the University of Minnesota claims to show that glyphosate—the main active ingredient in the widely used herbicide Roundup—is safe for humans to use. But guess whose business school just got a huge donation from Monsanto, the company that produces Roundup? That's right, the University of Minnesota. Ever hear of conflict of interest? This study is junk, just like the product it's defending.

This is a fallacy. It doesn't follow from the fact that the University received a grant from Monsanto that scientists working at that school faked the results of a study. But the fact of the grant does raise a red flag. There may be some conflict of interest at play. Such things have happened in the past (e.g., studies funded by Big Tobacco showing that smoking is harmless). But raising the possibility of a conflict is not enough, on its own, to show that the study in question can be dismissed out of hand. It may be appropriate to subject it to heightened scrutiny, but we cannot shirk our duty to assess its arguments on their merits.

A similar thing happens when we point to the *hypocrisy* of someone making a certain argument—when their actions are inconsistent with the conclusion they're trying to convince us of. Consider the following:

The head of the local branch of the American Federation of Teachers union wrote an op-ed yesterday in which she defended public school teachers from criticism and made the case that public schools' quality has never been higher. But guess what? She sends her own kids to private schools out in the suburbs! What a hypocrite. The public school system is a wreck and we need more accountability for teachers.

This passage makes a strong point, but then commits a fallacy. It would appear that, indeed, the AFT leader is hypocritical; her choice to send her kids to private schools suggests (but doesn't necessarily prove) that she doesn't believe her own assertions about the quality of public schools. Again, this raises a red flag about her arguments; it's a reason to subject them to heightened scrutiny. But it is not a sufficient reason to reject them out of hand, and to accept the opposite of her conclusions. That's committing a fallacy. She may have perfectly good reasons, having nothing to do with the allegedly low quality of public schools, for sending her kids to the private school in the suburbs. Or she may not. She may secretly think, deep down, that her kids would be better off not going to public schools. But none of this means her arguments in the op-ed should be dismissed; it's beside the point. Do her premises back up her conclusion? Are her premises true? That's how we evaluate an argument; hypocrisy on the part of the arguer doesn't relieve us of the responsibility to conduct thorough, dispassionate logical analysis.

A very specific version of the circumstantial *ad hominem*, one that involves pointing out one's opponent's hypocrisy, is worth highlighting, since it happens so frequently. It has its own Latin name: *tu quoque*, which translates roughly as "you, too." This is the "I know you are but what am I?" fallacy; the "pot calling the kettle black"; "look who's talking." It's a technique used in very specific circumstances: your opponent accuses you of doing or advocating something that's wrong, and, instead of making an argument to defend the rightness of your actions, you simply throw the accusation back in your opponent's face—they did it too. But that doesn't make it right!

An example. In February 2016, Supreme Court Justice Antonin Scalia died unexpectedly. President Obama, as is his constitutional duty, nominated a successor. The Senate is supposed to 'advise and consent' (or not consent) to such nominations, but instead of holding hearings on the nominee (Merrick Garland), the Republican leaders of the Senate

discussion grows longer, the probability of a comparison involving Nazis or Hitler approaches one." ("Meme, Counter-meme", Wired, 10/1/94)

declared that they wouldn't even consider the nomination. Since the presidential primary season had already begun, they reasoned, they should wait until the voters has spoken and allow the new president to make a nomination. Democrats objected strenuously, arguing that the Republicans were shirking their constitutional duty. The response was classic *tu quoque*. A conservative writer asked, "Does any sentient human being believe that if the Democrats had the Senate majority in the final year of a conservative president's second term—and Justice [Ruth Bader] Ginsburg's seat came open—they would approve any nominee from that president?"<sup>6</sup> Senate Majority Leader Mitch McConnell said that he was merely following the "Biden Rule," a principle advocated by Vice President Joe Biden when he was a Senator, back in the election year of 1992, that then-President Bush should wait until after the election season was over before appointing a new Justice (the rule was hypothetical; there was no Supreme Court vacancy at the time).

This is a fallacious argument. Whether or not Democrats would do the same thing if the circumstances were reversed is irrelevant to determining whether that's the right, constitutional thing to do.

The final variant of the circumstantial *ad hominem* fallacy is perhaps the most egregious. It's certainly the most ambitious: it's a preemptive attack on one's opponent to the effect that, because of the type of person he is, nothing he says on a particular topic can be taken seriously; he is excluded entirely from debate. It's called *poisoning the well*. This phrase was coined by the famous 19<sup>th</sup> century Catholic intellectual John Henry Cardinal Newman, who was a victim of the tactic. In the course of a dispute he was having with the famous Protestant intellectual Charles Kingsley, Kingsley is said to have remarked that anything Newman said was suspect, since, as a Catholic priest, his first allegiance was not to the truth (but rather to the Pope). As Newman rightly pointed out, this remark, if taken seriously, has the effect of rendering it impossible for him or any other Catholic to participate in any debate whatsoever. He accused Kingsley of "poisoning the wells."

We poison the well when we exclude someone from a debate because of who they are. Imagine an Englishman saying something like, "It seems to me that you Americans should reform your healthcare system. Costs over here are much higher than they are in England. And you have millions of people who don't even have access to healthcare. In the UK, we have the NHS (National Health Service); medical care is a basic right of every citizen." Suppose an American responded by saying, "What you know about it, Limey? Go back to England." That would be poisoning the well (with a little name-calling thrown in). The Englishman is excluded from debating American healthcare just because of who he is—an Englishman, not an American.

### III. Fallacies of Weak Induction

As their name suggests, what these fallacies have in common is that they are bad—that is, weak— inductive arguments. Recall, inductive arguments attempt to provide premises that make their conclusions more probable. We evaluate them according to how probable their conclusions are in light of their premises: the more probable the conclusion (given the premises), the stronger the argument; the less probable, the weaker. The fallacies of weak induction are arguments whose premises do not make their conclusions very probable—but that are nevertheless often successful in convincing people of their conclusions. We will discuss five informal fallacies that fall under this heading.

#### *Argument from Ignorance (Argumentum ad Ignorantiam)*

This is a particularly egregious and perverse fallacy. In essence, it's an inference from premises to the effect that there's a lack of knowledge about some topic to a definite conclusion about that topic. We don't know; therefore, we know!

Of course, put that baldly, it's plainly absurd; actual instances are more subtle. The fallacy comes in a variety of closely

6. David French, National Review, 2/14/16

related forms. It will be helpful to state them in bald/absurd schematic fashion first, then elucidate with more subtle real-life examples.

The first form can be put like this:

Nobody knows how to explain phenomenon X.

/∴ My crazy theory about X is true.

That sounds silly, but consider an example: those “documentary” programs on cable TV about aliens. You know, the ones where they suggest that extraterrestrials built the pyramids or something (there are books and websites, too). How do they get you to believe that crazy theory? By creating mystery! By pointing to facts that nobody can explain. The Great Pyramid at Giza is aligned (almost) exactly with the magnetic north pole! On the day of the summer solstice, the sun sets exactly between two of the pyramids! The height of the Great Pyramid is (almost) exactly one one-millionth the distance from the Earth to the Sun! How could the ancient Egyptians have such sophisticated astronomical and geometrical knowledge? Why did the Egyptians, careful recordkeepers in (most) other respects, (apparently) not keep detailed records of the construction of the pyramids? Nobody knows. Conclusion: aliens built the pyramids.

In other words, there are all sorts of (sort of) surprising facts about the pyramids, and nobody knows how to explain them. From these premises, which establish only our ignorance, we’re encouraged to conclude that we know something: aliens built the pyramids. That’s quite a leap— too much of a leap. The *burden of proof* falls upon those making the claims to justify them. It is not okay to present claims and insist that they be taken as true and until proven false. Moreover, those who do try to shift the burden of proof often suddenly become very rigorous and demanding of other’s attempts to refute their claims.

Another form this fallacy takes can be put crudely thus:

Nobody can PROVE that I’m wrong. \_\_\_\_\_

/∴ I’m right.

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

An example will help. There are lots of people who claim that evolutionary biology is a lie: there’s no such thing as evolution by natural selection, and it’s especially false to claim that humans evolved from earlier species, that we share a common ancestor with apes. Rather, the story goes, the Bible is literally true: the Earth is only about 6,000 years old, and humans were created as-is by God just as the Book of Genesis describes. The Argument from Ignorance is one of the favored techniques of proponents of this view. They are especially fond of pointing to “gaps” in the fossil record—the so-called “missing link” between humans and a pre-human, ape-like species—and claim that the incompleteness of the fossil record vindicates their position.

But this argument is an instance of the fallacy. The standard of proof—a complete fossil record without any gaps—is impossibly high. Evolution has been going on for a LONG time (the Earth is actually about 4.5 billion years old, and living things have been around for at least 3.5 billion years). So many species have appeared and disappeared over time that it’s absurd to think that we could even come close to collecting fossilized remains of anything but the tiniest fraction of them. It’s hard to become a fossil, after all: a creature has to die under special circumstances to even have a chance for its remains to do anything than turn into compost. And we haven’t been searching for fossils in a systematic way for very long (only since the mid-1800s or so). It’s no surprise that there are gaps in the fossil record, then. What’s surprising, in fact, is that we have as rich a fossil record as we do. Many, many transitional species have been discovered, both between humans and their ape-like ancestors, and between other modern species and their distant forbears (whales used to be land-based creatures, for example; we know this (in part) from the fossils of early protowhale species with longer and longer rear hip- and leg-bones).

We will never have a fossil record complete enough to satisfy skeptics of evolution. But their standard is unreasonably high, so their argument is fallacious. Sometimes they put it even more simply: nobody was around to witness evolution in action; therefore, it didn’t happen. This is patently absurd, but it follows the same pattern: an unreasonable standard of proof (witnesses to evolution in action; impossible, since it takes place over such a long period of time), followed by the leap to the unwarranted conclusion.

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

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

Of course lack of imagination on the part of an individual isn't evidence for or against a proposition, but people often argue this way. A (hilarious) example comes from the rap duo Insane Clown Posse in their 2009 single, "Miracles". Here's the line:

Water, fire, air and dirt  
F\*\*king magnets, how do they work?  
And I don't wanna talk to a scientist  
Y'all mother\*\*kers lying, and getting me pissed.

Violent J and Shaggy 2 Dope can't understand how there could be a scientific, non-miraculous explanation for the workings of magnets. They conclude, therefore, that magnets are miraculous.

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

No evidence has been found that X is true.  
/∴ X is false.

You may have heard the slogan, "Absence of evidence is not evidence of absence." This is an attempt to sum up this version of the fallacy. But it's not quite right. What it should say is that absence of evidence is not always *definitive* evidence of absence. An example will help illustrate the idea. During the 2016 presidential campaign, a reporter (David Fahrenthold) took to Twitter to announce that despite having "spent weeks looking for proof that [Donald Trump] really does give millions of his own [money] to charity..." he could only find one donation, to the NYC Police Athletic League. Trump has claimed to have given millions of dollars to charities over the years. Does this reporter's failure to find evidence of such giving prove that Trump's claims about his charitable donations are false? No. To rely only on this reporter's testimony to draw such a conclusion would be to commit the fallacy.

However, the failure to uncover evidence of charitable giving does provide *some* reason to suspect Trump's claims may be false. How much of a reason depends on the reporter's methods and credibility, among other things. But sometimes a lack of evidence can provide strong support for a negative conclusion. This is an inductive argument; it can be weak or strong. For example, despite multiple claims over many years (centuries, if some sources can be believed), no evidence has been found that there's a sea monster living in Loch Ness in Scotland. Given the size of the body of water, and the extensiveness of the searches, this is pretty good evidence that there's no such creature—a strong inductive argument to that conclusion. To claim otherwise—that there is such a monster, despite the lack of evidence—would be to commit the version of the fallacy whereby one argues "You can't PROVE I'm wrong; therefore, I'm right," where the standard of proof is unreasonably high.

One final note on this fallacy: it's common for people to mislabel certain bad arguments as arguments from ignorance; namely, arguments made by people who obviously don't know what the heck they're talking about. People who are confused or ignorant about the subject on which they're offering an opinion are liable to make bad arguments, but the fact of their ignorance is not enough to label those arguments as instances of the fallacy. We reserve that designation for arguments that take the forms canvassed above: those that rely on ignorance—and not just that of the arguer, but of the audience as well—as a premise to support the conclusion.

### *Appeal to Inappropriate Authority*

One way of making an inductive argument—of lending more credence to your conclusion—is to point to the fact that some relevant authority figure agrees with you. In law, for example, this kind of argument is indispensable: appeal to precedent (Supreme Court rulings, etc.) is the attorney's bread and butter. And in other contexts, this kind of move can make for a strong inductive argument. If I'm trying to convince you that fluoridated drinking water is safe and beneficial,

I can point to the Centers for Disease Control, where a wealth of information supporting that claim can be found.<sup>7</sup> Those people are scientists and doctors who study this stuff for a living; they know what they're talking about.

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

Usually, the inappropriateness of the authority being appealed to is obvious. But sometimes it isn't. A particularly subtle example is AstraZeneca's hiring of Dr. Phil McGraw in 2016 as a spokesperson for their diabetes outreach campaign. AstraZeneca is a drug manufacturing company. They make a diabetes drug called Bydureon. The aim of the outreach campaign, ostensibly, is to increase awareness among the public about diabetes; but of course the real aim is to sell more Bydureon. A celebrity like Dr. Phil can help. Is he an appropriate authority? That's a hard question to answer. It's true that Dr. Phil had suffered from diabetes himself for 25 years, and that he personally takes the medication. So that's a mark in his favor, authority-wise. But is that enough? We'll talk about how feeble Phil's sort of anecdotal evidence is in supporting general claims (in this case, about a drug's effectiveness) when we discuss the hasty generalization fallacy; suffice it to say, one person's positive experience doesn't prove that the drug is effective. But, Dr. Phil isn't just a person who suffers from diabetes; he's a doctor! It's right there in his name (everybody always simply refers to him as 'Dr. Phil'). Surely that makes him an appropriate authority on the question of drug effectiveness. Or maybe not. Phil McGraw is not a medical doctor; he's a PhD. He has a doctorate in Psychology. He's not a licensed psychologist; he cannot legally prescribe medication. He has no relevant professional expertise about drugs and their effectiveness. He is not an appropriate authority in this case. He looks like one, though, which makes this a very sneaky, but effective, advertising campaign.

### *Post hoc ergo propter hoc*

Here's another fallacy for which people always use the Latin, usually shortening it to 'post hoc'. The whole phrase translates to 'After this, therefore because of this', which is a pretty good summation of the pattern of reasoning involved. Crudely and schematically, it looks like this:

X occurred before Y.

/ ∴ X caused Y.

This is not a good inductive argument. That one event occurred before another gives you *some* reason to believe it might be the cause—after all, X can't cause Y if it happened after Y did—but not nearly enough to conclude that it is the cause. A silly example: I, your humble author, was born on June 19<sup>th</sup>, 1974; this was just shortly before a momentous historical event, Richard Nixon's resignation of the Presidency on August 9<sup>th</sup> later that summer. My birth occurred before Nixon's resignation; but this is (obviously!) not a reason to think that it caused his resignation.

Though this kind of reasoning is obviously shoddy—a mere temporal relationship clearly does not imply a causal relationship—it is used surprisingly often. In 2012, New York Yankees shortstop Derek Jeter broke his ankle. It just so happened that this event occurred immediately after another event, as Donald Trump pointed out on Twitter: "Derek Jeter broke ankle one day after he sold his apartment in Trump World Tower." Trump followed up: "Derek Jeter had a great career until 3 days ago when he sold his apartment at Trump World Tower- I told him not to sell- karma?" No, Donald, not karma; just bad luck.

Nowhere is this fallacy more in evidence than in our evaluation of the performance of presidents of the United States.

7. Check it out: <https://www.cdc.gov/fluoridation/>

Everything that happens during or immediately after their administrations tends to be pinned on them. But presidents aren't all-powerful; they don't cause everything that happens during their presidencies. On July 9<sup>th</sup>, 2016, a short piece appeared in the Washington Post with the headline "Police are safer under Obama than they have been in decades". What does a president have to do with the safety of cops? Very little, especially compared to other factors like poverty, crime rates, policing practices, rates of gun ownership, etc., etc., etc. To be fair, the article was aiming to counter the equally fallacious claims that increased violence against police was somehow caused by Obama. Another example: in October 2015, US News & World Report published an article asking (and purporting to answer) the question, "Which Presidents Have Been Best for the Economy?" It had charts listing GDP growth during each administration since Eisenhower. But while presidents and their policies might have some effect on economic growth, their influence is certainly swamped by other factors. Similar claims on behalf of state governors are even more absurd. At the 2016 Republican National Convention, Governors Scott Walker and Mike Pence—of Wisconsin and Indiana, respectively—both pointed to record-high employment in their states as vindication of their conservative, Republican policies. But some other states were also experiencing record-high employment at the time: California, Minnesota, New Hampshire, New York, Washington. Yes, they were all controlled by Democrats. Maybe there's a separate cause for those strong jobs numbers in differently governed states? Possibly it has something to do with the improving economy and overall health of the job market in the whole country? Holding rituals every day before dawn doesn't make the sun rise.

### *Slippery Slope*

Like the *post hoc* fallacy, the slippery slope fallacy is a weak inductive argument to a conclusion about causation. This fallacy involves making an insufficiently supported claim that a certain action or event will set off an unstoppable causal chain-reaction—putting us on a slippery slope—leading to some disastrous effect.

This style of argument was a favorite tactic of religious conservatives who opposed gay marriage. They claimed that legalizing same-sex marriage would put the nation on a slippery slope to disaster. Famous Christian leader Pat Robertson, on his television program *The 700 Club*, puts the case nicely. When asked about gay marriage, he responded with this:

We haven't taken this to its ultimate conclusion. You've got polygamy out there. How can we rule that polygamy is illegal when you say that homosexual marriage is legal? What is it about polygamy that's different? Well, polygamy was outlawed because it was considered immoral according to Biblical standards. But if we take Biblical standards away in homosexuality, well what about the other? And what about bestiality? And ultimately what about child molestation and pedophilia? How can we criminalize these things, at the same time have Constitutional amendments allowing same-sex marriage among homosexuals? You mark my words, this is just the beginning of a long downward slide in relation to all the things that we consider to be abhorrent.

This is a classic slippery slope fallacy; he even uses the phrase 'long downward slide'. The claim is that allowing gay marriage will force us to decriminalize polygamy, bestiality, child molestation, pedophilia—and ultimately, "all the things that we consider to be abhorrent." Yikes! That's a lot of things. Apparently, gay marriage will lead to utter anarchy.

There are genuine slippery slopes out there—unstoppable causal chain-reactions. But this isn't one of them. The mark of the slippery slope fallacy is the assertion that the chain can't be stopped, with reasons that are insufficient to back up that assertion. In this case, Pat Robertson has given us the abandonment of "Biblical standards" as the lubrication for the slippery slope. But this is obviously insufficient. Biblical standards are expressly forbidden, by the "establishment clause" of the First Amendment to the U.S. Constitution, from forming the basis of the legal code. The slope is not slippery. As recent history has shown, the legalization of same sex marriage does not lead to the acceptance of bestiality and pedophilia; the argument is fallacious.

Fallacious slippery slope arguments have long been deployed to resist social change. Those opposed to the abolition of slavery warned of economic collapse and social chaos. Those who opposed women's suffrage asserted that it would lead to the dissolution of the family, rampant sexual promiscuity, and social anarchy. Of course none of these dire predictions came true; the slopes simply weren't slippery.



## *Hasty Generalization*

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

One commits the hasty generalization fallacy when one makes this kind of inference based on an insufficient number of particular premises, when one is too quick—hasty—in inferring the general conclusion.

People who deny that global warming is a genuine phenomenon often commit this fallacy. In February of 2015, the weather was unusually cold in Washington, DC. Senator James Inhofe of Oklahoma famously took to the Senate floor wielding a snowball. “In case we have forgotten, because we keep hearing that 2014 has been the warmest year on record, I ask the chair, ‘You know what this is?’ It's a snowball, from outside here. So it's very, very cold out. Very unseasonable.” He then tossed the snowball at his colleague, Senator Bill Cassidy of Louisiana, who was presiding over the debate, saying, “Catch this.”

Senator Inhofe commits the hasty generalization fallacy. He's trying to establish a general conclusion—that 2014 wasn't the warmest year on record, or that global warming isn't really happening (he's on the record that he considers it a “hoax”). But the evidence he presents is insufficient to support such a claim. His evidence is an unseasonable coldness in a single place on the planet, on a single day. We can't derive from that any conclusions about what's happening, temperature-wise, on the entire planet, over a long period of time. That the earth is warming is not a claim that everywhere, at every time, it will always be warmer than it was; the claim is that, on average, across the globe, temperatures are rising. This is compatible with a couple of cold snaps in the nation's capital.

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

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

## IV. Fallacies of Illicit Presumption

This is a family of fallacies whose common characteristic is that they (often tacitly, implicitly) presume the truth of some claim that they're not entitled to. They are arguments with a premise (again, often hidden) that is assumed to be true, but is actually a controversial claim, which at best requires support that's not provided, which at worst is simply false. We will look at six fallacies under this heading.

### *Accident*

This fallacy is the reverse of the hasty generalization. That was a fallacious inference from insufficient particular

premises to a general conclusion; accident is a fallacious inference from a general premise to a particular conclusion. What makes it fallacious is an illicit presumption: the general rule in the premise is assumed, incorrectly, not to have any exceptions; the particular conclusion fallaciously inferred is one of the exceptional cases.

Here's a simple example to help make that clear:

Cutting people with knives is illegal.

Surgeons cut people with knives.

∴ Surgeons should be arrested.

One of the premises is the general claim that cutting people with knives is illegal. While this is true in almost all cases, there are exceptions—surgery among them. We pay surgeons lots of money to cut people with knives! It is therefore fallacious to conclude that surgeons should be arrested, since they are an exception to the general rule. The inference only goes through if we presume, incorrectly, that the rule is exceptionless.

Another example. Suppose I volunteer at my first grade daughter's school; I go in to her class one day to read a book aloud to the children. As I'm sitting down on the floor with the kiddies, crisscross applesauce, as they say, I realize that I can't comfortably sit that way because of the .44 Magnum revolver that I have tucked into my waistband. So I remove the piece from my pants and set it down on the floor in front of me, among the circled-up children. The teacher screams and calls the office, the police are summoned, and I'm arrested. As they're hauling me out of the room, I protest: "The Second Amendment to the Constitution guarantees my right to keep and bear arms! This state has a 'concealed carry' law, and I have a license to carry that gun! Let me go!"

I'm committing the fallacy of Accident in this story. True, the Second Amendment guarantees the right to keep and bear arms; but that rule is not without exceptions. Similarly, concealed carry laws also have exceptions—among them being a prohibition on carrying weapons into elementary schools. My insistence on being released only makes sense if we presume, incorrectly, that the legal rules I'm citing are without exception.

One more example from real life. After the financial crisis in 2008, the Federal Reserve—the central bank in the United States, whose task it is to create conditions leading to full employment and moderate inflation—found itself in a bind. The economy was in a free-fall, and unemployment rates were skyrocketing, but the usual tool it used to mitigate such problems—cutting the short-term federal funds rate (an interest rate banks charge each other for overnight loans)—was unavailable, because they had already cut the rate to zero (the lowest it could go). So they had to resort to unconventional monetary policies, among them something called "quantitative easing". This involved the purchase, by the Federal Reserve, of financial assets like mortgage-backed securities and longer-term government debt (Treasury notes).

Now, the nice thing about being the Federal Reserve is that when you want to buy something—in this case a bunch of financial assets—it's really easy to pay for it: you have the power to create new money out of thin air! That's what the Federal Reserve does; it controls the amount of money that exists. So if the Fed wants to buy, say, \$10 million worth of securities from Bank of America, they just press a button and presto—\$10 million dollars that didn't exist a second ago comes into being as an asset of Bank of America.

This quantitative easing policy was controversial. Many people worried that it would lead to runaway inflation. Generally speaking, the more money there is, the less each bit of it is worth. So creating more money makes things cost more—inflation. The Fed was creating money on a very large scale—on the order of a trillion dollars. Shouldn't that lead to a huge amount of inflation?

Economist Art Laffer thought so. In June of 2009, he wrote an op-ed in the *Wall Street Journal* warning that "[t]he unprecedented expansion of the money supply could make the '70s look benign."<sup>8</sup> (There was a lot of inflation in the '70s.)

Another famous economist, Paul Krugman, accused Laffer of committing the fallacy of accident. While it's generally true that an increase in the supply of money leads to inflation, that rule is not without exceptions. He had described such

8. Art Laffer, "Get Ready for Inflation and Higher Interest Rates," June 11, 2009, *Wall Street Journal*

exceptional circumstances in 1998,<sup>9</sup> and pointed out that the economy of 2009 was in that condition (which economists call a “liquidity trap”): “Let me add, for the 1.6 trillionth time, *we are in a liquidity trap*. And in such circumstances a rise in the monetary base does not lead to inflation.”<sup>10</sup>

It turns out Krugman was correct. The expansion of the monetary supply did not lead to runaway inflation; as a matter of fact, inflation remained *below* the level that the Federal Reserve wanted, barely moving at all. Laffer had indeed committed the fallacy of accident.

### *Begging the Question (Petitio Principii)*

First things first: ‘begging the question’ is not synonymous with ‘raising the question’; this is an extremely common usage, but it is wrong. You might hear a newscaster say, “Today Donald Trump’s private jet was spotted at the Indianapolis airport, which begs the question: ‘Will he choose Indiana Governor Mike Pence as running mate?’” This is a mistaken usage of ‘begs the question’; the newscaster should have said ‘raises the question’ instead.

‘Begging the question’ is a translation of the Latin *‘petitio principii’*, which refers to the practice of asking (begging, petitioning) your audience to grant you the truth of a claim (principle) as a premise in an argument—but it turns out that the claim you’re asking for is either identical to, or presupposes the truth of, the very conclusion of the argument you’re trying to make.

In other words, when you beg the question, you’re arguing in a circle: one of the reasons for believing the conclusion is the conclusion itself! It’s a Fallacy of Illicit Presumption where the proposition being presumed is the very proposition you’re trying to demonstrate; that’s clearly an illicit presumption.

Here’s a stark example. If I’m trying to convince you that Goldfinger is a dangerous idiot (the conclusion of my argument is ‘Goldfinger is a dangerous idiot’), then I can’t ask you to grant me the claim ‘Goldfinger is a dangerous idiot’. The premise can’t be the same as the conclusion. Imagine a conversation:

Me: “Goldfinger is a dangerous idiot.”

You: “Really? Why do you say that?”

Me: “Because Goldfinger is a dangerous idiot.”

You: “So you said. But why should I agree with you? Give me some reasons.”

Me: “Here’s a reason: Goldfinger is a dangerous idiot.”

And round and round we go. Circular reasoning; begging the question.

It’s not always so blatant. Sometimes the premise is not identical to the conclusion, but merely presupposes its truth. Why should we believe that the Bible is true? Because it says so right there in the Bible that it’s the infallible Word of God. This premise is not the same as the conclusion, but it can only support the conclusion if we take the Bible’s word for its own truthfulness, i.e., if we assume that the Bible is true. But that was the very claim we were trying to prove!

Sometimes the premise is just a re-wording of the conclusion. Consider this argument: “To allow every man unbounded freedom of speech must always be, on the whole, advantageous to the state; for it is highly conducive to the interests of the community that each individual should enjoy a liberty, perfectly unlimited, of expressing his sentiments.”<sup>11</sup> Replacing synonyms with synonyms, this comes down to “Free speech is good for society because free speech is good for society.” Not a good argument.

9. “But if current prices are not downwardly flexible, and the public expects price stability in the long run, the economy cannot get the expected inflation it needs; and in that situation the economy finds itself in a slump against which shortrun monetary expansion, no matter how large, is ineffective.” From Paul Krugman, “It’s baack: Japan’s Slump and the Return of the Liquidity Trap,” 1998, Brookings Papers on Economic Activity, 2

10. Paul Krugman, June 13, 2009, The New York Times

11. This is a classic example, from Richard Whately’s 1826 *Elements of Logic*.

## Loaded Questions

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

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

Strategic deployment of loaded yes-or-no questions can be an extremely effective debating technique. If you catch your opponent off-guard, they will struggle to respond to your question, since a simple ‘yes’ or ‘no’ commits them to the truth of the illicit presumption, which they want to deny. This makes them look evasive, shifty. And as they struggle to come up with a response, you can pounce on them: “It’s a simple question. Yes or no? Why won’t you answer the question?” It’s a great way to appear to be winning a debate, even if you don’t have a good argument. Imagine the following dialogue:

Liberal TV Host: “Are you or are you not in favor of the president’s plan to force wealthy business owners to pay their fair share in taxes to protect the vulnerable and aid this nation’s underprivileged?”

Conservative Guest: “Well, I don’t agree with the way you’ve laid out the question. As a matter of fact...”

Host: “It’s a simple question. Should business owners pay their fair share; yes or no?”

Guest: “You’re implying that the president’s plan would correct some injustice. But corporate taxes are already very...”

Host: “Stop avoiding the question! It’s a simple yes or no!”

Combine this with the sort of subconscious appeal to force discussed above—yelling, fingerpointing, etc.—and the host might come off looking like the winner of the debate, with his opponent appearing evasive, uncooperative, and inarticulate.

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

A famous example of this occurred during the Republican presidential primary in 2000. George W. Bush was the front-runner, but was facing a surprisingly strong challenge from the upstart John McCain. After McCain won the New Hampshire primary, he had a lot of momentum. The next state to vote was South Carolina; it was very important for the Bush campaign to defeat McCain there and reclaim the momentum. So they conducted a push poll designed to spread negative feelings about McCain—by implanting false beliefs among the voting public. “Pollsters” called voters and asked, “Would you be more or less likely to vote for John McCain for president if you knew he had fathered an illegitimate black child?” The aim, of course, is for voters to come to believe that McCain fathered an illegitimate black child. But he did no such thing. He and his wife *adopted* a daughter, Bridget, from Bangladesh.

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

## *False Choice*

This fallacy occurs when someone tries to convince you of something by presenting it as one of limited number of options and the best choice among those options. The illicit presumption is that the options are limited in the way presented; in fact, there are additional options that are not offered. The choice you're asked to make is a false choice, since not all the possibilities have been presented. Other times, the offered choice is between two options which are presented as being mutually exclusive when they are not.

## *Composition*

The fallacy of Composition rests on an illicit presumption about the relationship between a whole thing and the parts that make it up. This is an intuitive distinction, between whole and parts: for example, a person can be considered as a whole individual thing; it is made up of lots of parts— hands, feet, brain, lungs, etc., etc. We commit the fallacy of Composition when we mistakenly assume that any property that all of the parts share is also a property of the whole. Schematically, it looks like this:

All of the parts of X have property P.

Any property shared by all of the parts of a thing is also a property of the whole.

$\therefore$  X has the property P.

The second premise is the illicit presumption that makes this argument go through. It is illicit because it is simply false: sometimes all the parts of something have a property in common, but the whole does not have that property.

Consider the 1980 U.S. Men's Hockey Team. They won the gold medal at the Olympics that year, beating the unstoppable-seeming Russian team in the semifinals. (That game is often referred to as "The Miracle on Ice" after announcer Al Michaels' memorable call as the seconds ticked off at the end: "Do you believe in miracles? Yes!") Famously, the U.S. team that year was a rag-tag collection of no-name college guys; the average age on the team was 21, making them the youngest team ever to compete for the U.S. in the Olympics. The Russian team, on the other hand, was packed with seasoned hockey veterans with world-class talent.

In this example, the team is the whole, and the individual players on the team are the parts. It's safe to say that one of the properties that all of the parts shared was mediocrity—at least, by the standards of international competition at the time. They were all good hockey players, of course— Division I college athletes—but compared to the Hall of Famers the Russians had, they were mediocre at best. So, all of the parts have the property of being mediocre. But it would be a mistake to conclude that the whole made up of those parts—the 1980 U.S. Men's Hockey Team—also had that property. The team was *not* mediocre; they defeated the Russians and won the gold medal! They were a classic example of the whole being greater than the sum of its parts.

## *Division*

The fallacy of Division is the exact reverse of the fallacy of Composition. It's an inference from the fact that a whole has some property to a conclusion that a part of that whole has the same property, based on the illicit presumption that wholes and parts must have the same properties. Schematically:

X has the property P.

Any property of a whole thing is shared by all of its parts.

$\therefore$  x, which is a part of X, has property P.

The second premise is the illicit presumption. It is false, because sometimes parts of things don't have the same properties as the whole. George Clooney is handsome; does it follow that his large intestine is also handsome? Of course not. Toy Story 3 is a funny movie. Remember when Mr. Potato Head had to use a tortilla for his body? Or when Buzz gets

flipped into Spanish mode and does the flamenco dance with Jessie? Hilarious. But not all of the parts of the movie are funny. When it looks like all the toys are about to be incinerated at the dump? When Andy finally drives off to college? Not funny at all!

## V. Fallacies of Linguistic Emphasis

Natural languages like English are unruly things. They're full of ambiguity, shades of meaning, vague expressions; they grow and develop and change over time, often in unpredictable ways, at the capricious collective whim of the people using them. Languages are messy, complicated. This state of affairs can be taken advantage of by the clever debater, exploiting the vagaries of language to make convincing arguments that are nevertheless fallacious. This exploitation involves the manipulation of linguistic forms to *emphasize* facts, claims, emotions, etc. that favor one's position, and to de-emphasize those that do not. We will survey four techniques that fall under this heading.

### *Accent*

This is one of the original 13 fallacies that Aristotle recognized in his *Sophistical Refutations*. Our usage, however, will depart from Aristotle's. He identifies a potential for ambiguity and misunderstanding that is peculiar to his language—ancient Greek. That language—in written form—used diacritical marks along with the alphabet, and transposition of these could lead to changes in meaning. English is not like this, but we can identify a fallacy that is roughly in line with the spirit of Aristotle's accent: it is possible, in both written and spoken English (along with every other language), to convey different meanings by stressing individual words and phrases. The devious use of stress to emphasize contents that are helpful to one's rhetorical goals, and to suppress or obscure those that are not—that is the fallacy of accent.

There are a number of techniques one can use with the written word that fall in the category of accent. Perhaps the simplest way to emphasize favorable contents, and de-emphasize unfavorable ones, is to vary the size of one's text. We see this in advertising all the time. You drive past a store that's having a sale, which they advertise with a sign in the window. In the largest, most eye-catching font, you read, "70% OFF!" "Wow," you might think, "that's a really steep discount. I should go in to the store and get a great deal." At least, that's what the store wants you to think. They're emphasizing the fact of (at least one) steep discount. If you look more closely at the sign, however, you'll see the things that they're legally required to say, but that they'd like to deemphasize. There's a tiny 'Up to' in front of the gigantic '70% OFF!'. For all you know, there's one crappy item that nobody wants, tucked in the back of the store, that's discounted at 70%; everything else has much smaller discounts, or none at all. Also, if you squint really hard, you'll see an asterisk after the '70% OFF!'; which leads to some text at the bottom of the poster, in the tiniest font possible, that reads, "While supplies last. See store details. Not available in all locations. Offer not valid weekends or holidays. All sales are final." This is the proverbial "fine print". It makes the sale look a lot less exciting. So they hide it.

Footnotes are generally a good place to hide unfavorable content. We all know that CEOs of big companies—especially banks—get paid ridiculous sums of money. Some of it is just their salary and stock options; those amounts are huge enough to turn most people off. But there are other perks that are so over-the-top, companies and executives feel like it's best to hide them from the public (and their shareholders) in the footnotes of CEO contracts and SEC reports. Michelle Leder runs a website called [footnoted.com](http://footnoted.com), which is dedicated to combing through these documents and exposing outrageous compensation packages. She's uncovered executives spending over \$700,000 to renovate their offices, demanding helicopters in addition to their corporate jets, receiving millions of dollars' worth of private security services, etc., etc. These additional, extravagant forms of compensation seem excessive to most people, so companies do all they can to hide them from the public.

Another abuse of footnotes can occur in academic or legal writing. Legal briefs and opinions and academic papers

seek to persuade. If you're writing such a document, and you relegate a strong objection to your conclusion to a brief mention in the footnotes, you're de-emphasizing that point of view and making it less likely that the reader will reject your arguments. That's a fallacious suppression of opposing content, a sneaky trick to try to convince people you're right without giving them a forthright presentation of the merits (and demerits) of your position.

The fallacy of accent can occur in speech as well as writing. The audible correlate of "fine print" is that guy talking really fast at the end of the commercial, rattling off all the unpleasant side effects and legal disclaimers that, if given a full, deliberate presentation might make you less likely to buy the product they're selling. The reason, by the way, that we know about such horrors as the possibility of driving while not awake (a side-effect of some sleep aids) and a four-hour erection (side-effect of erectile-dysfunction drugs), is that drug companies are required, by federal law, *not* to commit the fallacy of accent if they want to market drugs directly to consumers. They have to read what's called a "major statement" that lists all of these side-effects explicitly, and no fair cramming them in at the end and talking over them really fast.

When we speak, how we stress individual words and phrases can alter the meaning that we convey with our utterances. Consider the sentence 'These pretzels are making me thirsty.' Now consider various utterances of that sentence, each stressing a different word; different meanings will be conveyed:

These pretzels are making me thirsty. [Not those over there, these right here.]

These *pretzels* are making me thirsty. [It's not the chips, it's the pretzels.]

These pretzels *are* making me thirsty. [Don't try to tell me they're not; they are.]

And so on. We can capture the various stresses typographically by using italics (or boldface or all-caps), but if we leave that out, we lose some of the meaning conveyed by the actual, stressed utterance. One can commit the fallacy of accent by transcribing someone's speech in a way that omits stress-indicators, and thereby obscures or alters the meaning that the person actually conveyed. Suppose a candidate for president says, "I *hope* this country never has to wage war with Iran." The stress on 'hope' clearly conveys that the speaker doubts that his hopes will be realized; the candidate has expressed a suspicion that there may be war with Iran. This speech might set off a scandal: saying such a thing during an election could negatively affect the campaign, with the candidate being perceived as a war-monger; it could upset international relations. The campaign might try to limit the damage by writing an op-ed in a major newspaper, and transcribing the candidate's utterance without any indication of stress: "The Senator said, 'I hope this country never has to wage war with Iran.' This is a sentiment shared by most voters, and even our opponent." This transcription, of course, obscures the meaning of the original utterance. Without the stress, there is not additional implication that the candidate suspects that there will in fact be a war.

## *Quoting out of Context*

Another way to obscure or alter the meaning of what someone actually said is to quote them selectively. Remarks taken out of their proper context might convey a different meaning than they did within that context.

Consider a simple example: movie ads. These often feature quotes from film critics, which are intended to convey the impression that the movie was well-liked by them. "Critics call the film 'unrelenting', 'amazing', and 'a one-of-a-kind movie experience'", the ad might say. That sounds like pretty high praise. I think I'd like to see that movie. That is, until I read the actual review from which those quotes were pulled:

I thought I'd seen it all at the movies, but even this jaded reviewer has to admit that this film is something new, a one-of-a-kind movie experience: two straight hours of unrelenting, snooze-inducing mediocrity. I find it amazing that not one single aspect of this movie achieves even the level of "eh, I guess that was OK."

The words 'unrelenting' and 'amazing'—and the phrase 'a one-of-a-kind movie experience'—do in fact appear in that review. But situated in their original context, they're doing something completely different than the movie ad would like us to believe.

Politicians often quote each other out of context to make their opponents look bad. In the 2012 presidential campaign,

both sides did it rather memorably. The Romney campaign was trying to paint President Obama as anti-business. In a campaign speech, Obama once said the following:

If you've been successful, you didn't get there on your own. You didn't get there on your own. I'm always struck by people who think, well, it must be because I was just so smart. There are a lot of smart people out there. It must be because I worked harder than everybody else. Let me tell you something: there are a whole bunch of hardworking people out there. If you've got a business, you didn't build that. Somebody else made that happen.

Yikes! What an insult to all the hard-working small-business owners out there. They didn't build their own businesses? The Romney campaign made some effective ads, with these remarks playing in the background, and small-business people describing how they struggled to get their firms going. The problem is, that quote above leaves some bits out—specifically, a few sentences before the last two. Here's the full transcript:

If you've been successful, you didn't get there on your own. You didn't get there on your own. I'm always struck by people who think, well, it must be because I was just so smart. There are a lot of smart people out there. It must be because I worked harder than everybody else. Let me tell you something: there are a whole bunch of hardworking people out there.

If you were successful, somebody along the line gave you some help. There was a great teacher somewhere in your life. Somebody helped to create this unbelievable American system that we have that allowed you to thrive. Somebody invested in roads and bridges. If you've got a business, you didn't build that. Somebody else made that happen.

Oh. He's not telling business owners that they didn't build their own businesses. The word 'that' in "you didn't build that" doesn't refer to the businesses; it refers to the roads and bridges—the "unbelievable American system" that makes it possible for businesses to thrive. He's making a case for infrastructure and education investment; he's not demonizing small-business owners.

The Obama campaign pulled a similar trick on Romney. They were trying to portray Romney as an out-of-touch billionaire, someone who doesn't know what it's like to struggle, and someone who made his fortune by buying up companies and firing their employees. During one speech, Romney said: "I like being able to fire people who provide services to me." Yikes! What a creep. This guy gets off on firing people? What, he just finds joy in making people suffer? Sounds like a moral monster. Until you see the whole speech:

I want individuals to have their own insurance. That means the insurance company will have an incentive to keep you healthy. It also means if you don't like what they do, you can fire them. I like being able to fire people who provide services to me. You know, if someone doesn't give me the good service that I need, I want to say I'm going to go get someone else to provide that service to me.

He's making a case for a particular health insurance policy: self-ownership rather than employer-provided health insurance. The idea seems to be that under such a system, service will improve since people will be empowered to switch companies when they're dissatisfied—kind of like with cell phones, for example. When he says he likes being able to fire people, he's talking about being a savvy consumer. I guess he's not a moral monster after all.

## *Equivocation*

Typical of natural languages is the phenomenon of homonymy: when words have the same spelling and pronunciation, but different meanings—like 'bat' (referring to the nocturnal flying mammal) and 'bat' (referring to the thing you hit a baseball with). This kind of natural-language messiness allows for potential fallacious exploitation: a sneaky debater can manipulate the subtleties of meaning to convince people of things that aren't true—or at least not justified based on what they say. We call this kind of maneuver the fallacy of equivocation

Here's an example. Consider a banker; let's call him Fred. Fred is the president of a bank, a real big-shot. He's married, but he's not faithful: he's carrying on an affair with one of the tellers at his bank, Linda. Fred and Linda have a favorite activity: they take long lunches away from their workplace, having romantic picnics at a beautiful spot they found a short



walk away. They lay out their blanket underneath an old, magnificent oak tree, which is situated right next to a river, and enjoy champagne and strawberries while canoodling and watching the boats float by.

One day—let’s say it’s the anniversary of when they started their affair—Fred and Linda decide to celebrate by skipping out of work entirely, spending the whole day at their favorite picnic spot. (Remember, Fred’s the boss, so he can get away with this.) When Fred arrives home that night, his wife is waiting for him. She suspects that something is up: “What are you hiding, Fred? Are you having an affair? I called your office twice, and your secretary said you were ‘unavailable’ both times. Tell me this: Did you even go to work today?” Fred replies, “Scout’s honor, dear. I swear I spent all day at the bank today.”

See what he did there? ‘Bank’ can refer either to a financial institution or the side of a river—a river bank. Fred and Linda’s favorite picnic spot is on a river bank, and Fred did indeed spend the whole day at *that* bank. He’s trying to convince his wife he hasn’t been cheating on her, and he exploits this little quirk of language to do so. That’s equivocation.

A similar linguistic phenomenon can also be exploited to equivocate: polysemy. This is distinct from, but similar to, homonymy. The meanings of homonyms are typically unrelated. In polysemy, the same word or phrase has multiple, related meanings—different senses. Consider the word ‘law’. The meaning that comes immediately to mind is the statutory one: “A rule of conduct imposed by authority.”<sup>12</sup> The state law prohibiting murder is an instance of a law in this sense. There is another sense of ‘law’, however; this is the sense operative when we speak of scientific laws. These are regularities in nature—Newton’s law of universal gravitation, for example. These meanings are similar, but distinct: statutes, human laws, are prescriptive; scientific laws are descriptive. Human laws tell us how we *ought* to behave; scientific laws *describe* how things actually do, and *must*, behave. Human laws can be violated: I could murder someone. Scientific laws cannot be violated: if two bodies have mass, they will be attracted to one another by a force directly proportional to the product of their masses and inversely proportional to the square of the distance between them; there’s no getting around it.

A common argument for the existence of God relies on equivocation between these two senses of ‘law’:

There are laws of nature.

By definition, laws are rules imposed by an Authority.

So the laws of nature were imposed by an Authority.

The only Authority who could impose such laws is an all-powerful Creator—God.

∴ God exists.

This argument relies on fallaciously equivocating between the two senses of ‘law’—human and natural. It’s true that human laws are by definition imposed by an authority; but that is not true of natural laws. Additional argument is needed to establish that those must be so imposed.

A famous instance of equivocation of this sort occurred in 1998, when President Bill Clinton denied having an affair with White House intern Monica Lewinsky by declaring forcefully in a press conference: “I did not have sexual relations with that woman—Ms. Lewinsky.” The president wanted to convince his audience that nothing sexually inappropriate had happened, even though, as was revealed later, lots of icky sex stuff had been going on. He does this by taking advantage of the polysemy of the phrase ‘sexual relations’. In the broadest sense, the phrase connotes sexual activity of any kind—including oral sex (which Bill and Monica engaged in). This is the sense the president wants his audience to have in mind, so that they’re convinced by his denial that nothing untoward happened. But a more restrictive sense of ‘sexual relations’—a bit more old-fashioned and Southern usage—refers specifically to intercourse (which Bill and Monica did not engage in). It’s this sense that the president can fall back on if anyone accuses him of having lied; he can claim that, strictly speaking, he was telling the truth: he and Monica didn’t have ‘relations’ in the intercourse sense. Clinton later admitted to “misleading” the American people—but, importantly, not to lying.

The distinction between lying and misleading is a hard one to draw precisely, but roughly speaking it’s the difference between trying to get someone to believe something false by saying something false (lying) and trying to get them

12. From the Oxford English Dictionary.

to believe something false by saying something true but deceptive (misleading). Besides homonymy and polysemy, yet another common linguistic phenomenon can be exploited to this end. This phenomenon is implicature, identified and named by the philosopher Paul Grice in the 1960s.<sup>13</sup> Implicatures are contents that we communicate over and above the literal meaning of what we say—aspects of what we *mean* by our utterances that aren't stated explicitly. People listening to us infer these additional meanings based on the assumption that the speaker is being cooperative, observing some unwritten rules of conversational practice. To use one of Grice's examples, suppose your car has run out of gas on the side of the road, and you stop me as I walk by, explaining your plight, and I say, "There's a gas station right around the corner." Part of what I communicate by my utterance is that the station is open and selling gas right now—that you can go there and solve your problem. You can infer this content based on the assumption that I'm being a cooperative conversational partner; if the station is closed or out of gas—and I knew it—then I would be acting unhelpfully, uncooperatively. Notice, though, that this content is not part of what I literally said: all I told you is that there is a gas station around the corner, which would still be true even if it were closed and/or out of gas.

Implicatures are yet another subtle aspect of meaning in natural language that can be exploited. So a final technique that we might classify under the fallacy of equivocation is false implication—saying things that are strictly speaking true, but which communicate false implicatures. Grocery stores do this all the time. You know those signs posted under, say, cans of soup that say "10 for \$10"? That's the store's way of telling us that soup's on sale for a buck a can; that's right, you don't need to buy 10 cans to get the deal; if you buy one can, it's \$1; 2 cans are \$2, and so on. So why not post a sign saying "\$1 per can"? Because the 10-for-\$10 sign conveys the false implicature that you need to buy 10 cans in order to get the sale price. The store's trying to drive up sales.

A striking example of false implicature is featured in one of the most prominent U.S. Supreme Court rulings on perjury law. In the original criminal case, a defendant by the name of Bronston had the following exchange with the prosecuting attorney: "Q. Do you have any bank accounts in Swiss Banks, Mr. Bronston? A. No, sir. Q. Have you ever? A. The company had an account there for about six months, in Zurich."<sup>14</sup> As it turns out, Bronston did not have any Swiss bank accounts at the time of the questioning, so his first answer was strictly true. But he did have Swiss bank accounts in the past. However, his second answer does not deny this. All he says is that his company had Swiss bank accounts—an answer that implicates that he himself did not. Based on this exchange, Bronston was convicted of perjury, but the Supreme Court overturned that conviction, pointing out that Bronston had not made any false statements (a requirement of the perjury statute); the falsehood he conveyed was an implicature.

### *Manipulative Framing*

Words are powerful. They can trigger emotional responses and activate associations with related ideas, altering the way we perceive the world and conceptualize issues. The language we use to describe a particular policy, for example, can affect how favorably our listeners are likely to view that proposal. How we *frame* issues with language can profoundly influence how persuasive our arguments about those issues will be. The technique of choosing words to frame issues intentionally to manipulate your audience is what we will call the fallacy of manipulative framing.

The importance of framing in politics has long been recognized, but only in recent decades has it been raised to an art form. One prominent practitioner of the art is Republican consultant Frank Luntz. In a 200-plus page memo he sent to Congressional Republicans in 1997, and later in a book, Luntz stressed the importance of choosing persuasive language to frame issues so that voters would be more likely to support Republican positions on issues.<sup>15</sup> One of his recommendations illustrates manipulative framing nicely. In the United States, if you leave a fortune to your heirs after

13. See his *Studies in the Way of Words*, 1989, Cambridge: Harvard University Press.

14. *Bronston v. United States*, 409 US 352 - Supreme Court 1973

15. Frank Luntz, 2007, *Words That Work: It's Not What You Say, It's What People Hear*. New York: Hyperion.

you die, then the government taxes it (provided it's greater than about \$5.5 million, or \$11 million for a couple, as of 2016). The usual name for this tax is the 'estate tax'. Luntz encouraged Republicans—who are generally opposed to this tax—to start referring to it instead as the “death tax”. This framing is likelier to cause voters to oppose the tax as well: taxing people for dying? Talk about kicking a man when he's down! (Polling bears this out: people oppose the tax in higher numbers when it's called the 'death tax' than when it's called the 'estate tax'.)

The linguist George Lakoff has written extensively on the subject of framing.<sup>16</sup> His remarks on the subject of “tax relief” nicely illustrate how framing works:

On the day that George W. Bush took office, the words tax relief started appearing in White House communiqués to the press and in official speeches and reports by conservatives. Let us look in detail at the framing evoked by this term.

The word relief evokes a frame in which there is a blameless Afflicted Person who we identify with and who has some Affliction, some pain or harm that is imposed by some external Cause-of-pain. Relief is the taking away of the pain or harm, and it is brought about by some Reliever-of-pain.

The Relief frame is an instance of a more general Rescue scenario, in which there a Hero (The Reliever-of-pain), a Victim (the Afflicted), a Crime (the Affliction), A Villain (the Cause-of-affliction), and a Rescue (the Pain Relief). The Hero is inherently good, the Villain is evil, and the Victim after the Rescue owes gratitude to the Hero.

The term tax relief evokes all of this and more. Taxes, in this phrase, are the Affliction (the Crime), proponents of taxes are the Causes-of Affliction (the Villains), the taxpayer is the Afflicted Victim, and the proponents of “tax relief” are the Heroes who deserve the taxpayers' gratitude.

Every time the phrase tax relief is used and heard or read by millions of people, the more this view of taxation as an affliction and conservatives as heroes gets reinforced.<sup>17</sup> Carefully chosen words can trigger all sorts of mental associations, mostly at the subconscious level, that affect how people perceive the issues and have the power to change opinions. That's why manipulative framing is ubiquitous in public discourse.

Consider debates about illegal immigration. Those who are generally opposed to policies that favor such people will often refer to them as “illegal immigrants”. This framing emphasizes the fact that they are in this country illegally, making it likelier that the listener will also oppose policies that favor them. A further modification can further increase this likelihood: “illegal aliens.” The word ‘alien’ has a subtle dehumanizing effect; if we don't think of them as individual people with hopes and dreams, we're not likely to care much about them. Even more dehumanizing is a framing one often sees these days: referring to illegal immigrants simply as “illegals”. They are the living embodiment of illegality! Those who advocate on behalf of such people, of course, use different terminology to refer to them: “undocumented workers”, for example. This framing de-emphasizes the fact that they're here illegally; they're merely “undocumented”. They lack certain pieces of paper; what's the big deal? It also emphasizes the fact that they are working, which is likely to cause listeners to think of them more favorably.

The use of manipulative framing in the political sphere extends to the very names that politicians give the laws they pass. Consider the healthcare reform act passed in 2010. Its official name is The Patient Protection and Affordable Care Act. Protection of patients, affordability, care—these all trigger positive associations. The idea is that every time someone talks about the law prior to and after its passage, they will use the name with this positive framing and people will be more likely to support it. As you may know, this law is commonly referred to with a different moniker: ‘Obamacare’. This is the framing of choice for the law's opponents: any negative associations people have with President Obama are attached to the law; and any negative feelings they have about healthcare reform get attached to Obama. Late night talk show host Jimmy Kimmel demonstrated the effectiveness of framing on his show one night in 2013. He sent a crew outside his studio to interview people on the street and ask them which approach to health reform they preferred, the Affordable Care Act or Obamacare. Overwhelmingly, people expressed a preference for the Affordable Care Act

16. See. e.g. Lakoff, G. 2004, *Don't Think of an Elephant!* Chelsea Green Publishing.

17. George Lakoff, 2/14/2006, “Simple Framing,” Rockridge Institute.

over Obamacare, even though those are just two different ways of referring to the same piece of legislation. Framing is especially important when the public is ignorant of the actual content of policy proposals, which is all too often the case.

## EXERCISES

Identify the fallacy most clearly exhibited in the following passages.

1. Responding to a critical comment from one Mike Helgeson, the anonymous proprietor of the “Governor Scott Walker Sucks” Facebook page wrote this:

“Mike Helgeson is a typical right wing idiot who assumes anyone who doesn’t like Walker doesn’t work and is living off the government. I work 60-70 hours a week during the summer so get a clue and quit whining like a child.”

2. Randy: “I think abortion should be illegal. Unborn children have a right not to be killed.”

Sally: “What do you know about it? You’re a man.”

3. We need a balanced budget amendment, forcing the U.S. government to balance its budget every year. All of the states have to balance their budgets; so should the country.

4. Privacy is important to the development of full individuals because there has to be an interior zone within each person that other people don’t see.<sup>18</sup>

5. Of course, the real gripe the left has in Wisconsin is that the current legislative districts were drawn by Republicans, who were granted that right due to their large victories in 2010. Since the new maps were unveiled in 2011, Democrats have issued several legal challenges trying to argue the maps are “unfair” and that Republicans overstepped their bounds.

Did Republicans draw the maps to their advantage? Of course they did — just as Democrats would have done had they held control of state government in 2010.<sup>19</sup>

6. President Obama has been terrible for healthcare costs in this county. When we had our first child, before he was president, we only paid a couple of hundred dollars out of pocket; insurance covered the rest. The new baby we just had? The hospital bills cost us over \$5,000!

7. Let’s call our public schools what they really are—‘government’ schools.<sup>20</sup>

8. You shouldn’t hire that guy. The last company he worked for went bankrupt. He’s probably a failure, too.

9. Fred: “I read about a new study that shows diet soda is good for weight loss—better than water, even.”

Fiona: “Yeah, but look at who sponsored it: the International Life Sciences Institute, which is a non-profit, but whose board of directors is stacked with people from Coca-Cola and PepsiCo.”

10. Buy the Amazing RonCo Super Bass-o-Matic 76, the easiest way to prepare delicious bass: only 3 installments of \$19.99.\*

\*Shipping and handling fees apply. Price is before state, local, and federal taxes. Safety goggles sold separately. The rush from using Super Basso-Matic 76 has been shown to be addictive in laboratory mice. RonCo not legally responsible for injury or choking death due to ingestion of insufficiently pureed bass. The following aquatic species cannot be safely prepared using the Super Bass-o-Matic: shark, cod, squid, octopus, catfish, dogfish, crab (snow, blue, and king), salmon, tuna, lobster, crayfish, crawfish, crawdaddy, whale (sperm, killer, and humpback). Super Bass-o-Matic is illegal to operate in the following jurisdictions: California, Massachusetts, Canada, the European Union, Haiti.

11. Former pro golfer Brandel Chamblee, expressing concern about the workout habits of current pro Rory McIlroy: “I think of what happened to Tiger Woods. And I think more than anything of what Tiger Woods did early in his career

18. David Brooks, 4/14/15, New York Times

19. Christian Schneider, 7/14/16, Milwaukee Journal-Sentinel

20. John Stossel, 10/2/13, foxnews.com

with his game was just an example of how good a human being can be, what he did towards the middle and end of his career is an example to be wary of. That's just my opinion. And it does give me a little concern when I see the extensive weightlifting that Rory is doing in the gym."

12. Former pro golfer Gary Player, famous for his rigorous workouts and long career, responding on McIlroy's behalf via Twitter:

"Haha, too funny. Don't worry about the naysayers mate. They all said I would be done at 30 too."

13. Responding to North Korean rhetoric about pre-emptive nuclear strikes if S. Korea and U.S. engage in war games, Russia:

"We consider it to be absolutely impermissible to make public statements containing threats to deliver some 'preventive nuclear strikes' against opponents," said the statement, as translated by the Russian TASS news agency. "Pyongyang should be aware of the fact that in this way the DPRK [North Korea] will become fully opposed to the international community and will create international legal grounds for using military force against itself in accordance with the right of a state to self-defense enshrined in the United Nations Charter."<sup>21</sup>

14. Responding to criticism that the state university system was declining in quality under his watch due to a lack of funding, the Governor said, "Look, we can either have huge tuition increases, which no one wants, or university administrators and professors can learn to do more with less."

15. Man, I told you flossing was useless. Look at this newspaper article, "Medical benefits of flossing unproven":

"The two leading professional groups — the American Dental Association and the American Academy of Periodontology, for specialists in gum disease and implants — cited other studies as proof of their claims that flossing prevents buildup of gunk known as plaque, early gum inflammation called gingivitis, and tooth decay. However, most of these studies used outdated methods or tested few people. Some lasted only two weeks, far too brief for a cavity or dental disease to develop. One tested 25 people after only a single use of floss. Such research, like the reviewed studies, focused on warning signs like bleeding and inflammation, barely dealing with gum disease or cavities."<sup>22</sup>

17. Did you hear about Jason Pierre-Paul, the defensive end for the New York Giants? He blew off half his hand lighting off fireworks on the Fourth of July. Man, jocks are such idiots.

18. Mother of recent law school grad, on the phone with her son: "Did you pass the bar?"

Son: "Yes, mom."

[He failed the bar exam. But he did walk past a tavern on his way home from work.]

19. Alfred: "I'm telling you, Obama is a socialist. He said, and I quote, 'I actually believe in redistribution.'"

Betty: "C'mon. Read the whole interview, he said: 'I think the trick is figuring out how do we structure government systems that pool resources and hence facilitate some redistribution because I actually believe in redistribution, at least at a certain level to make sure that everybody's got a shot. How do we pool resources at the same time as we decentralize delivery systems in ways that both foster competition, can work in the marketplace, and can foster innovation at the local level and can be tailored to particular communities.' Socialists don't talk about 'decentralization,' 'competition,' and 'the marketplace.' That's straight-up capitalism."

20. In 2016, Supreme Court Justice Ruth Bader Ginsburg gave an interview in which she criticized Republican presidential candidate Donald Trump, calling him a "faker" and saying she couldn't imagine him as president. She was criticized for these remarks: as a judge, she's supposed to be politically impartial, the argument went; her remarks amounted to a violation of judicial ethics. Defenders of Ginsburg were quick to point out that her late colleague, Justice Antonin Scalia, was a very outspoken conservative on a variety of political issues, and even went hunting with Vice President Dick Cheney one time before he was set to hear a case in which Cheney was involved. Isn't that a violation of judicial ethics, too?

21. Horace: "Man, these long lines at the airport are ridiculous. No liquids on the plane, taking off my shoes, full-body scans. Is all this really necessary?"

21. 3/18/16, The Daily Caller

22. Jeff Donn, 8/2/16, "Medical benefits of dental floss unproven," Associated Press

Pete: “Of course it is. TSA precautions prevent terrorism. There hasn’t been a successful terrorist attack in America involving planes since these extra security measures went into place, has there?”

22. A married couple goes out to dinner, and they have a bit too much wine to drink. After some discussion, they decide nevertheless to drive home. Since the wife is the more intoxicated of the two, the husband takes the wheel. On the way home, he’s pulled over by the police. When asked whether he’s had anything to drink that night, he replies, with a nod toward his wife, “She did. That’s why I’m driving.”

## Deductive logic I: Aristotelian Logic

### I. Deductive Logic

In this and the next 3 chapters we will study deductive logic and two approaches to evaluating deductive arguments. The first, which is the subject of the present chapter, was developed by Aristotle nearly 2,500 years ago, and we'll refer to it simply as Aristotelian Logic; the second, the subject of the next chapter, has roots nearly as ancient as Aristotle's but wasn't fully developed until the 19<sup>th</sup> century, and is called Propositional Logic.

Again, these are two different approaches to the same problem: evaluating deductive arguments, determining whether they are valid or invalid. Recall, deductive arguments are valid just in case their premises guarantee their conclusions; and validity is determined entirely by the *form* of the argument. The two logics we study will have different ways of identifying the logical form of arguments, and different methods of testing those forms for validity. These are two of the things a deductive logic must do: specify precise criteria for determining logical form and develop a way of testing it for validity.

But before a logic can do those two things, there is a preliminary job: it must tame natural language. Real arguments that we care about evaluating are expressed in natural languages like English, Greek, etc. As we saw in our discussion of the logical fallacies in the last chapter, natural languages are unruly: they are filled with ambiguity and vagueness, and exhibit an overall lack of precision that makes it very difficult to conduct the kind of rigorous analysis necessary to determine whether or not an argument is valid. So before making that determination, a logic must do some tidying up; it must remove the imprecision inherent in natural language expressions of arguments and make them suitable for rigorous analysis. There are various approaches to this task. Aristotelian Logic and Sentential logic adopt two different strategies.

Aristotelian Logic seeks to tame natural language by restricting itself to a well-behaved, precise *portion* of the language. It only evaluates arguments that are expressed within that precisely delimited subset of the language. Propositional logic achieves precision by eschewing natural language entirely: it constructs its own artificial language, and only evaluates arguments expressed in its terms.

This strategy may seem overly restrictive: if we limit ourselves to arguments expressed in a limited vocabulary—and especially if we leave behind natural language entirely—aren't we going to miss lots of (all?) arguments that we care about? The answer is no, these approaches are not nearly as restrictive as they might seem. We can *translate* back and forth between the special portion of language in Aristotelian Logic and expressions in natural language that are outside its scope. Likewise, we can translate back and forth between the artificial language of Sentential logic and natural language. The process of translating *from* the unruly bits of natural language *into* these more precise alternatives is what removes the ambiguity, vagueness, etc. that stands in the way of rigorous analysis and evaluation. So, part of the task of taming natural language is showing how one's alternative to it is nevertheless related to it—how it picks out the logically important features of natural language arguments while leaving behind their extraneous, recalcitrant bits.

These, then, are the three tasks that a deductive logic must accomplish:

1. Tame natural language.
2. Precisely define logical form.
3. Develop a way to test logical forms for validity.

The process for evaluating real arguments expressed in natural language is to render them precise and suitable for evaluation by translating them into the preferred vocabulary developed in step 1, then to identify and evaluate their forms according to the prescriptions of steps 2 and 3.

We now proceed to discuss Aristotelian Logic, starting with its approach to taming natural language.

## II. Classes and Categorical Propositions

For Aristotle, the fundamental logical unit is the *class*. Classes are just sets of things—sets that we can pick out using language. The simplest way to identify a class is by using a plural noun: trees, clouds, asteroids, people—these are all classes. Names for classes can be grammatically more complex, too. We can modify the plural noun with an adjective: ‘rich people’ picks out a class. Prepositional phrases can further specify: ‘rich people from Italy’ picks out a different class. The modifications can go on indefinitely: ‘rich people from Italy who made their fortunes in real estate and whose grandmothers were rumored to be secret lovers of Benito Mussolini’ picks out yet another class—which is either very small, or possibly empty, I don’t know. (Empty classes are just classes with no members; we’ll talk more about them later.) We will refer to names of classes as ‘class-terms’, or just ‘terms’ for short. Since for Aristotle the fundamental logical unit is the class, and since terms are the bits of language that pick out classes, Aristotle’s logic is often referred to as a ‘term logic’. This is in contrast to the method of logic we will study in the next chapter, Propositional Logic, so-called because it takes the fundamental logical unit to be the proposition.

Of course, Aristotelian Logic must also deal with propositions—we’re evaluating arguments here, and by definition those are just sets of propositions—but since classes are the fundamental logical unit, Aristotle restricts himself to a particular kind of proposition: *categorical propositions*. ‘Category’ is just a synonym of ‘class’. Categorical propositions are propositions that make a claim about the relationship between two classes. This is the first step in taming natural language: Aristotelian Logic will only evaluate arguments made up entirely of categorical propositions. The specific type of arguments it deals with are called *syllogisms*. We’re limiting ourselves to a restricted portion of language—sentences expressing these kinds of propositions, which will feature two class terms—terms picking out the classes whose relationship is described in the categorical proposition. Soon, we will place further restrictions on the forms these sentences can take, but for now we will discuss categorical propositions generally.

Again, categorical propositions make an assertion about the relationship between two classes. There are three possibilities here:

- (1) Whole Inclusion: one class is contained entirely within the other.

Example: Class 1 = people; Class 2 = bipeds. The first class is entirely contained in the second; every person is a biped.

- (2) Partial Inclusion: one class is partially contained within the other; the two classes have at least one member in common.

Example: Class 1 = people; Class 2 = swimmers. Some people swim; some don’t. Some swimmers are people; some aren’t (fish, e.g.). These two classes overlap, but not entirely.

- (3) Exclusion: the two classes don’t have any members in common; they are exclusive. Example: Class 1 = people; Class 2 = birds. No people are birds; no birds are people. Batman notwithstanding (dude’s not really a bat; also, bats aren’t birds; robins are, but again, Robin’s not actually a bird, just a guy who dresses up like one).

Given these considerations, we can (more or less) formally define categorical propositions:

A categorical *proposition* is a claim about the relationship between two classes—call them S and P — that either affirms or denies that S is wholly or partially included in P.

Aristotle noted that, given this definition, there are four types of categorical proposition. We will discuss them in turn.



## *The Four Types of Categorical Proposition*

### Universal affirmative (A)

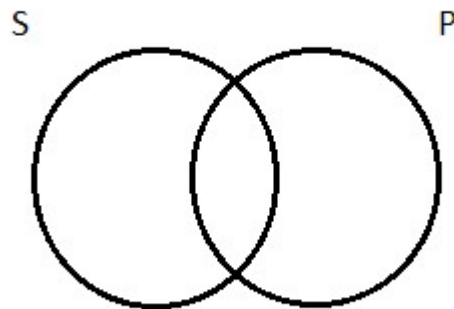
This type of proposition affirms the whole inclusion of the class S in the class P—it says that each member of S is also a member of P. The canonical expression of this proposition is a sentence of the form ‘All S are P’.

It is worth noting at this point why we chose ‘S’ and ‘P’ as the symbols for generic class terms. That’s because the former is the grammatical *subject* (S) of the sentence, and the latter is the grammatical *predicate* (P). This pattern will hold for the other three types of categorical proposition.

Back to the universal affirmative, A proposition. It affirms whole inclusion. For example, the sentence ‘All men are mortals’ expresses a proposition of this type, one that is true. ‘All men are Canadians’ also expresses a universal affirmative proposition, one that is false.

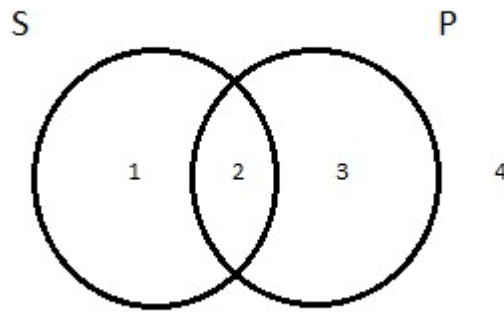
For the sake of concreteness, let’s choose subject and predicate classes that we can use as go-to examples as we talk about each of the four types of categorical proposition. Let’s let S = logicians and P = Germans. The A proposition featuring these two classes is expressed by ‘All logicians are Germans’. (We’ll remain agnostic about whether it’s true or false.)

When it comes time to test arguments for validity—the last step in the process we’ve just begun— it will be convenient for us to represent the four types of categorical propositions pictorially. The basic form of the pictures will be two overlapping circles, with the left-hand circle representing the subject class and the right-hand circle representing the predicate class. Like this:



To depict the four types of categorical propositions, we’ll modify this basic two-circle diagram by shading in parts of it or making marks inside the circles.

Before we get to the specific depiction of the A proposition, though, let’s talk about what the basic two-circle diagram does. It divides the universe into four regions, to which we can assign numbers like this:



Let's talk about what's inside each of the four regions if we take S to be the class of logicians and P to be the class of Germans.

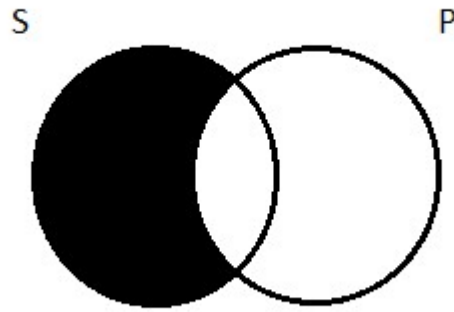
Region 1 is the portion of the S circle that doesn't overlap with the P circle. These are things in the subject class but outside the predicate class; they are logicians who aren't Germans. Aristotle was Greek, not German. So Aristotle is one of the residents of region 1—a logician who's not a German.

Region 2 is the area of overlap between the subject and predicate classes; its residents are members of both. So here we have the logicians who are also Germans. Gottlob Frege, a 19<sup>th</sup> century German logician, is the most important innovator in the history of logic other than Aristotle. He was German. So Frege lives in region 2; he's both a logician and a German.

Region 3 is the portion of the P circle that doesn't overlap with S. These are members of the predicate class—Germans, in our example—who are *not* members of the subject class—not logicians. This is where the non-logician Germans live. Wagner is a resident of region 3. The guy is clearly a German — but *not* a logician.

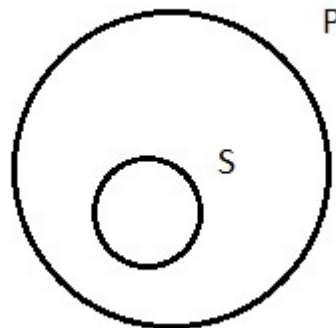
Region 4 is—everything else. It's all the things that are outside both the subject and predicate classes—things that are *neither logicians nor* Germans. You know who isn't German, but isn't a logician, either? Beyoncé. She lives in region 4. But so do lots and lots and lots of other things: the planet Jupiter is neither a logician nor a German; it's in there with Beyoncé, too. As is the left-front tire of my wife's car. And the second-smallest brick in the Great Wall of China. And so on.

So much for the blank two-circle diagram and how it carves up the universe. What we want to figure out is how to alter that diagram so that we end up with a picture of the universal affirmative proposition. Our particular example of an A proposition is that all logicians are Germans. How do we draw a picture of that, using the two circles as our starting point? Well, think about it this way: when we say all logicians are Germans, what we're really saying is that a certain kind of thing doesn't exist; there's no such thing as a non-German logician. In other words, despite what I said above about Aristotle, region 1 is empty, according to this proposition (which, again, may or may not be true; it doesn't matter whether it's true or not; we're just trying to figure out how to draw a picture that captures the claim it makes). To depict emptiness, we will adopt the convention of shading in the relevant region(s) of the diagram. So our picture of the universal affirmative looks like this:



All S are P means that you won't find any members of S that are outside the P circle (no logicians who aren't Germans). The place in the diagram where they might've been such things is blotted out to indicate its emptiness. The only portion of S that remains as a viable space is inside the P circle, in what we called region 2 (the logicians you do find will all be Germans).

A reasonable question could be raised at this point: why did we draw the universal affirmative that way, instead of another, alternative and possibly more intuitive way? A propositions affirm whole inclusion—that S is entirely contained within P. Isn't the obvious way to depict that state of affairs more like this:



S entirely contained within P. Easy. Why bother with the overlapping circles and the shading?

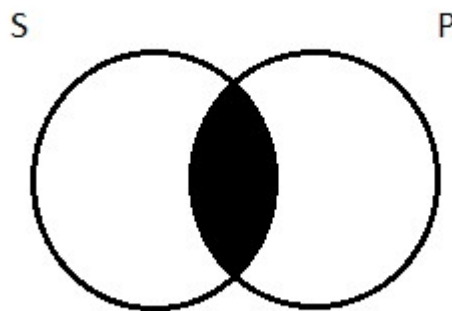
There's nothing wrong with this alternative depiction of the universal affirmative; it captures the claim being made. We adopt the first alternative depiction for purely practical reasons: when it comes time to test arguments for validity, we're going to use these pictures; and our method will depend on our four types of categorical propositions all being depicted starting with the same basic two-overlapping-circle diagram, with shading and marks inside. These diagrams, as you may know, are called "Venn Diagrams". They are named after the 19<sup>th</sup> century English logician John Venn, who invented them specifically as an easier means of testing arguments for validity in Aristotelian Logic (things were more unwieldy before Venn's innovation). It turns out Venn's method only works if we start with the overlapping circles for all four of the types of categorical proposition. So that's what we go with.

### Universal negative (E)

This type of proposition denies that S is even partially included in P. Put another way: it affirms that S and P are exclusive—that they have no members in common. The canonical expression of this proposition is a sentence of the

form 'No S are P'. So, for example, the sentence 'No dogs are cats' expresses a true universal negative proposition; the sentence 'No animals are cats' expresses a false one.

Again, we want to think about how to depict this type of proposition using the standard two-circle Venn diagram. Think about the proposition that no logicians are Germans. How do we draw a picture of this claim? Well, as we said, E propositions tell us that the two classes don't have any members in common. The region of the two-circle diagram where there are members of both classes is the area of overlap in the picture (what we referred to as region 2 above). The universal negative proposition tells us that there's nothing in there. So if I claim that no logicians are Germans, I'm saying that, contrary to my claims above about the German-ness of Gottlob Frege, no, there's no such thing as a logician-German. Region two is empty, and so we shade it out:



#### Particular affirmative (I)

This type of proposition affirms that S is partially included in P. Its canonical expression is a sentence of the form 'Some S are P'. So, for example, 'Some sailors are pirates' expresses a true particular affirmative proposition; 'Some sumo wrestlers are pigeons' expresses a false one.

Before we talk about how to depict I propositions with a Venn diagram, we need to discuss the word 'some'. Remember, in Aristotelian Logic we're taming natural language by restricting ourselves to a well-behaved portion of it—sentences expressing categorical propositions. We're proposing to use sentences with the word 'some' in them. 'Some', however, is not particularly well-behaved, and we're going to have to get it in line before we proceed.

Consider this utterance: "Some Republican voters are gun owners." This is true, and it communicates to the listener the fact that there's some overlap between the classes of Republican voters and gun owners. But it also communicates something more—namely, that some of those Republicans *aren't* gun owners. This is a fairly typical implicature (recall our discussion of this linguistic phenomenon in Chapter 2, when we looked at the fallacy of equivocation): when we say that some are, we also communicate that some are not.

But there are times when we use 'some' and don't implicate that some are not. Suppose you're talking to your mom, and you mention that you're reading a logic book. For some reason, your mom's always been curious about logic books, and asks you whether they're a good read. You respond, "Well, mom, I can tell you this for sure: *Some* logic books are boring. You should see this book I'm reading now; it's a total snooze-fest!" In this case, you say that some logic books are boring based on your experience with this particular book, but you *do not* implicate that some logic books are not boring; for all you know, all logic books are boring—it's just impossible to write an exciting logic book. This is a perfectly legitimate use of the word 'some', where all it means is that there is at least one: when you utter 'some logic books are boring', all you communicate is that there is at least one boring logic book.

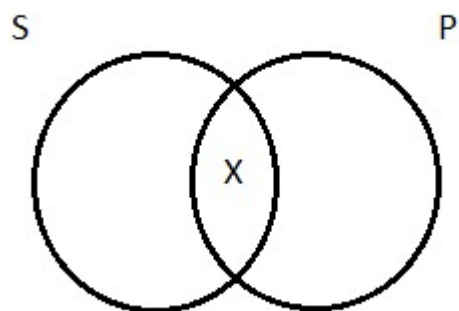
This is a bit of natural-language unruliness that we must deal with: sometimes when we use the word 'some', we implicate that some are not; other times, we don't, only communicating that at least one is. When we use 'some' in Aristotelian Logic, we need to know precisely what's being said. So we choose: 'some' means 'there is at least one'. 'Some S are P' tells us that those two classes have at least one member in common, and nothing more. 'Some sailors are pirates'

means that there's at least one sailor who's also a pirate, and that's it. There is no implication that some sailors are *not* pirates; at least one of them is, and for all we know, *all* of them are.

This can confuse people, so it's worth repeating. Heck, let's indent it:

‘Some’ means ‘there is at least one’, and that's it. It does not imply that some aren't.

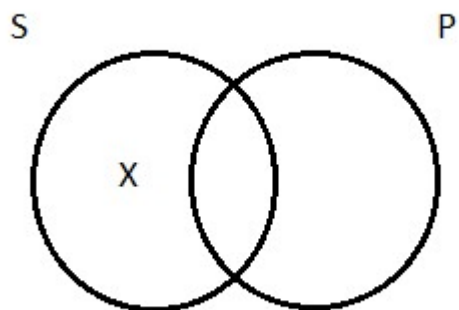
With that out of the way, we can turn our attention to the Venn diagram for the particular affirmative. It makes the assertion that S and P have at least one member in common. Turning to our concrete example, the sentence ‘Some logicians are Germans’ makes the claim that there is at least one logician who is a German. How do we draw a picture of this? We need to indicate that there's at least one thing in the area of overlap between the two circles on the diagram—at least one thing inside of region 2. We do this by drawing an X:



### Particular negative (O)

This type of proposition denies that S is wholly included in P. It claims that there is at least one member of S that is not a member of P. Given that ‘some’ means ‘there is at least one’, the canonical expression of this proposition is ‘Some S are not P’—there's at least one member of S that the two classes do not have in common. ‘Some sailors are not pirates’ expresses a true particular negative proposition; ‘Some dogs are not animals’ expresses a false one.

The Venn diagram for O propositions is simple. We need to indicate, on our picture, that there's at least one thing that's inside of S, but outside of P. To depict the fact that some logicians are not jerks, we need to put Aristotle (again, not a German, I'm pretty sure) inside the S circle, but outside the P circle. As with the diagram for the I proposition, we indicate the existence of at least one thing by drawing an X in the appropriate place:



## A Note on Terminology

It is commonly said that the four types of categorical propositions each have a quantity and a quality. There are two quantities: universal and particular. There are two qualities: affirmative and negative. There are four possible combinations of quantity and quality, hence four types of categorical proposition.

The universal propositions—A and E, affirmative and negative—are so-called because they each make a claim about the entire subject class. If I claim that all hobos are whiskey drinkers, I've made an assertion that covers every single hobo, every member of that class. Similarly, if I claim that no chickens are race car drivers, I've made an assertion covering all the chickens—they all fail to drive race cars.

The particular propositions—I and O, affirmative and negative—on the other hand, do *not* make claims about every member of the subject class. 'Some dinosaurs were herbivores' just makes the claim that there was at least one plant-eating dinosaur; we don't learn about all the dinosaurs. Similar remarks apply to an O proposition like 'Some dinosaurs were not carnivores'. Remember, 'some' just means 'at least one'.

The affirmative propositions—A and I, universal and particular—make affirmative claims about the relationship between two classes. A propositions affirm whole inclusion; I propositions affirm partial inclusion. Trivial fact: the Latin word meaning 'I affirm' is *affirmo*; the A and the I in that word are where the one-letter nicknames for the universal and particular affirmatives come from.

The negative propositions—E and O, universal and particular—make negative claims about the relationship between two classes. E propositions deny even partial inclusion; O propositions deny whole inclusion. Trivial fact: the Latin word meaning 'I deny' is *nego*; the E and the O in that word are where the one-letter nicknames for the universal and particular negatives come from.

### *Standard form for Sentences Expressing Categorical Propositions*

To tame natural language, Aristotelian Logic limits itself to that portion of language that expresses categorical propositions. Above, we gave "canonical" sentences for each of the four types of categorical proposition: 'All S are P' for the universal affirmative; 'No S are P' for the universal negative; 'Some S are P' for the particular affirmative; and 'Some S are not P' for the particular negative. These are not the only ways of expressing these propositions in English, but we will restrict ourselves to these standard forms. That is, we will only evaluate arguments whose premises and conclusion are expressed with sentences with these canonical forms.

Generally speaking, here is the template for sentences qualifying as standard form:

[Quantifier] Subject Term <copula> (not) Predicate Term

Standard form sentences begin with a quantifier—a word that indicates the quantity of the categorical proposition expressed. Restriction: only sentences beginning with 'All', 'No', or 'Some' qualify as standard form.

Subject and predicate terms pick out the two classes involved in the categorical proposition. Restriction: subject and predicate terms must be nouns or noun-phrases (nouns with modifiers) in order for a sentence to be in standard form.

The copula is a version of the verb 'to be' ('are', 'is', 'were', 'will be', etc.). Degree of freedom: it doesn't matter which version of the copula occurs in the sentence; it may be any number or tense. 'Some sailors are pirates' and 'Some sailors were pirates' both count as standard form, for example.

The word 'not' occurs in the standard form expression of the particular negative, O proposition: 'Some sailors are not pirates'. Restriction: the word 'not' can only occur in sentences expressing O propositions; 'not' appearing with any quantifier other than 'some' is a deviation from standard form.

We now have a precise delimitation of the portion of natural language to which Aristotelian Logic restricts itself: only sentences in standard form. But now a worry that we raised earlier becomes acute: if we can only evaluate arguments whose premises and conclusions are expressed with standard form sentences, aren't we *severely*, perhaps *ridiculously*, constrained? Has anyone, ever, outside a logic book, expressed a real-life argument that way?

This is where translation comes in. Lots of sentences that are not in standard form can be translated into standard form sentences that have the same meaning. Aristotle himself believed that all propositions, no matter how apparently complex or divergent, could ultimately be analyzed as one of the four types of categorical proposition. Though this is, to put it mildly, not a widely held belief today, it still had an enormous influence in the history of logic, since Aristotle's system was preeminent for more than 2,000 years. Over that time, logicians developed ever more elaborate procedures for analyzing a dizzying variety of non-Standard form sentences as expressing one of the four types of categorical propositions, and translating them accordingly. An exhaustive survey of those inquiries would be exhausting, and beyond the scope of this book. It will be enough to look at a few simple examples to get an idea of how many apparently deviant expressions can be treated by Aristotelian Logic. Our goal is simply to allay concerns that in restricting ourselves to standard form sentences we are severely limiting our logic's power to evaluate real-life arguments.

Let's consider the first deductively valid argument we encountered in this book, the one about Socrates: All men are mortal; Socrates is a man; therefore, Socrates is mortal. This argument has three propositions in it, but none of the three sentences expressing them are in standard form. The first sentence, 'All men are mortal', may appear to fit the bill, but it has a subtle flaw: 'mortal' is an adjective, not a noun. Class terms are required to be nouns or noun phrases. But this is an easy fix: add an 's' to the end and you get a plural noun. 'All men are mortals' is in standard form; it expresses a universal affirmative, A proposition. This prescription applies generally. Predicate adjectives can be replaced with suitable noun phrases most easily by just inserting the generic noun 'things': 'Some men are handsome' becomes 'Some men are handsome things'; 'No priests are silly' becomes 'No priests are silly things'.

Back to the Socrates argument. The second premise is also problematic: 'Socrates is a man'. First of all, it doesn't have a quantifier. Second, its subject term, 'Socrates', picks out an individual person; we're supposed to be dealing with classes here, right? Well, that's right, but it's not really a problem. We can just make the subject class a unit class—a class containing exactly one member, namely Socrates. Now we can understand the sentence as expressing the claim that the single member of that class is also a member of the class of men. That is, it's a universal affirmative—there's whole inclusion of the Socrates unit-class in the class of men. The sentence we need, then, starts with the quantifier 'All', and to make the grammar work, we pick a plural noun to name the Socrates class: 'All Socrateses are men'. Is 'Socrateses' the plural of 'Socrates'? I can't think of anything better. Anyway, the point is, that word picks out a class that has exactly one member, Socrates. Sentences with singular subjects can be rendered as universals. If I had the sentence 'Socrates is not alive', I could render it as a universal negative: 'No Socrateses are living things'.

There are other things to consider. English comes with a variety of quantifier words: 'each', 'every', 'any', and so on. Common sense tells us how to translate sentences featuring these into standard form: switch to the appropriate standard form quantifier—'All', 'No', or 'Some'. 'Every teacher is a hard worker' becomes 'All teachers are hard workers', for example. Sometimes quantifier words are omitted, but it's clear from context what's going on. 'Dogs are animals' means 'All dogs are animals'; 'People are waiting in line' can be rendered as 'Some people are things that are waiting in line'. Some sentences have a verb other than the copula. 'Some people eat rabbit', for example, can be translated into 'Some people are rabbit-eaters'. Sometimes the word 'not' appears in a sentence that has a quantifier other than 'some'. 'Not all mammals are carnivores', for example, can be translated into 'Some mammals are not carnivores'.

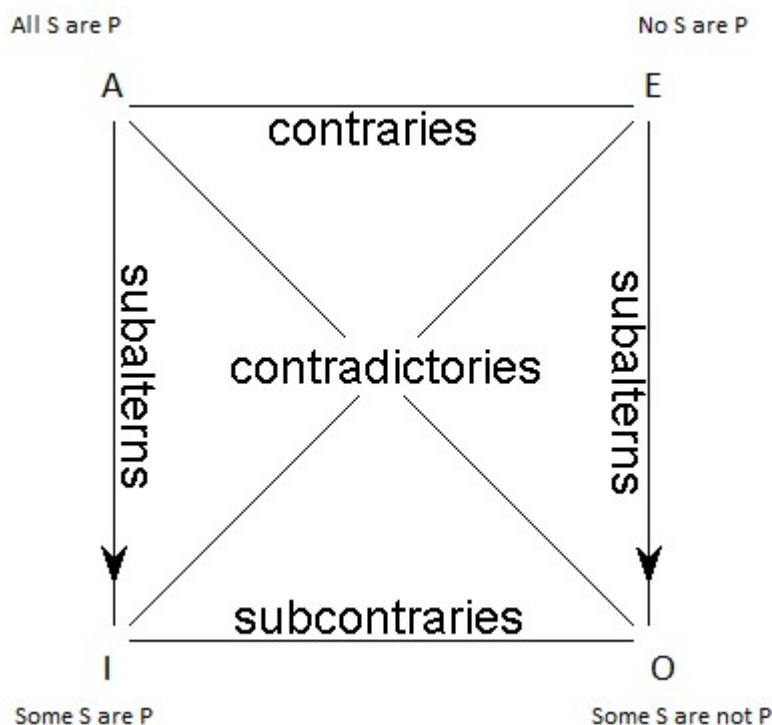
The list goes on. As I said, centuries of work has been done on the task of translating sentences into standard form. We can stop here, I think, and simply accept that the restriction to standard form sentences does not seriously limit the arguments that Aristotelian Logic can evaluate.

### III. The Square of Opposition

Having established the boundaries of our domain of logically well-behaved natural language, we turn now to an investigation of the properties of its inhabitants. The four types of categoricals are related to one another in systematic ways; we will look at those relationships.

The relationships are inferential: we can often infer, for example, from the truth of one of the four categoricals, whether

the other three are true or false. These inferential relationships among the four categorical propositions are summarized graphically in a diagram: The Square of Opposition. The diagram looks like this:



The four types of categorical propositions are arranged at the four corners of the square, and along the sides and diagonals are marked the relationships that obtain between pairs of them.

Now the Square of Opposition has a funny history. As we shall see, it involves some hidden commitments to what certain words, especially “some” entail. In fact, although Aristotle and the logical system that bears his name followed this square for nearly 2,000 years, there are now parts of the square to which we no longer adhere. Section V discusses these new changes and how such changes affect the types of inferences we can make with the square and categorical propositions, in general.

However, the classic Square of Opposition still provides a great way to explore potential relationships between categorical propositions in a graphic manner. Being able to follow along with these potential relationships, even if we adjust them down the road, can help us become more comfortable manipulating the ideas behind A, E, I, and O propositions. And so, it is to these relationships that we will now turn.

#### Contradictories

Contradictory pairs of categorical propositions are at opposite corners from one another on the Square of Opposition. A and O propositions are contradictory; E and I propositions are contradictory. What it means for a pair of propositions to be contradictory is this: they have opposite truth-values; when one is true, the other must be false, and *vice versa*.

This is pretty intuitive. Consider an A proposition—all sailors are pirates. Suppose I make that claim. How do you contradict me? How do you prove I’m wrong? “My brother’s in the Navy,” you might protest. “He’s a sailor, but he’s not a pirate.” That would do the trick. The way you contradict a universal affirmative claim—a claim that all S are P—is by showing that there’s at least one S (a sailor in this case, your brother) who’s not a P (not a pirate, as your brother is not). At least one S that’s not a P—that’s just the particular negative, O proposition, that some S are not P. (Remember: ‘some’ means ‘there is at least one’.) A and O propositions make opposite, contradictory claims. If it’s false that all sailors are pirates, then it must be true that some of them aren’t; that’s just how you show it’s false. Likewise, if it’s true that all dogs are animals (it is), then it must be false that some of them are not (you’re not going to find even one dog that’s not an animal). A and O propositions have opposite truth-values.

Likewise for E and I propositions. If I claim that no saints are priests, and you want to contradict me, what you need to do is come up with a saint who was a priest. It’s not hard: Saint Thomas Aquinas (who was the most prominent medieval



interpreter of Aristotle, by the way, and a terrific philosopher in his own right) was a priest. So, to contradict universal negative claim—that no S are P—you need to show that there's at least one S (a saint in this case, Thomas Aquinas) who is in fact a P (a priest, as Aquinas was). At least one S that is a P—that's just the particular affirmative, I proposition, that some S are P. (Again, 'some' means 'there is at least one'.) E and I propositions make opposite, contradictory claims. If it's false that no saints are priests, it must be true that some of them are; that's just how you show it's false. Likewise, it's true that no cats are dogs (it is), then it must be false that some of them are (you're not going to find even one cat that's a dog). E and I propositions have opposite truth-values.

## Contraries

The two universal propositions—A and E, along the top of the Square—are a *contrary* pair. This is a slightly weaker form of opposition than being *contradictory*. Being contrary means that they can't both be true, but they *could* both be false—though they needn't both be false; one could be true and the other false.

Again, this is intuitive. Suppose I claim the universal affirmative, "All dogs go to heaven," and you claim the corresponding universal negative, "No dogs go to heaven." (Those sentences aren't in standard form, but the translation is easy.) Obvious observation: we can't both be right; that is, both claims can't be true. On the other hand, we could both be wrong. Suppose getting into heaven, for dogs, is the way they say it is for people: if you're good and stuff, then you get in; but if you're bad, oh boy—it's the Other Place for you. In that case, both of our claims are false: some dogs (the good ones) go to heaven, but some dogs (the bad ones, the ones who bite kids, maybe) don't. But that picture might be wrong, too. I could be right and you could be wrong: God loves all dogs equally and they get a free pass into heaven. Or, I could be wrong and you could be right: God hates dogs and doesn't let any of them in; or maybe there is no heaven at all, and so nobody goes there, dogs included.

## Subcontraries

Along the bottom of the Square we have the two particular propositions—I and O—and they are said to be *subcontraries*. This means they can't both be false, but they *could* both be true—though they needn't be; one could be true and the other false.

It's easy to see how both I and O could be true. As a matter of fact, some sailors are pirates. That's true. Also, as a matter of fact, some of them are not. It's also easy to see how one of the particular propositions could be true and the other false, provided we keep in mind that 'some' just means 'there is at least one'. It's true that some dogs are mammals—that is, there is at least one dog that's a mammal—so that I proposition is true. In fact, all of them are—the A proposition is true as well. Which means, since A and O are contradictories, that the corresponding O proposition—that some dogs are not mammals—must be false. Likewise, it's true that some women are not (Catholic) priests (at least one woman isn't a priest); and it's false that some women are priests (the Church doesn't allow it). So O can be true while I is false.

It's a bit harder to see why both particular propositions can't be false. Why can't 'Some surfers are priests' and 'Some surfers are not priests' both be false? It's not immediately obvious. But think it through: if the I (some surfers are priests) is false, that means the E (no surfers are priests) must be true, since I and E are contradictory; and if the O (some surfers are not priests) is false, that means the A (all surfers are priests) must be true, since O and A are contradictory. That is to say, if I and O were both false, then the corresponding A and E propositions would both have to be true. But, as we've seen already, this is (obviously) impossible: if I claim that all surfers are priests and you claim that none of them are, we can't both be right.

## Subalterns

The particular propositions at the bottom of the table—I and O—are *subalterns* of the universal propositions directly above them—A and E, respectively.<sup>1</sup> This means that the pairs have the following relationship: if the universal proposition is true, then the particular proposition (its subaltern) must also be true. That is, if an A proposition is true, its corresponding I proposition must also be true; if an E proposition is true, its corresponding O proposition must also be true.

This is intuitive if we keep in mind, as always, that ‘some’ means ‘there is at least one’. Suppose the A proposition that all whales are mammals is true (it is); then the corresponding I proposition, that some whales are mammals, must also be true. Again, ‘some whales are mammals’ just means ‘at least one whale is a mammal’; if all of them are, then at least one of them is! Similarly, on the negative side of the square, if it’s true the no priests are women (universal negative, E), then it’s got to be true that some priests are not women (particular negative, O)—that at least one priest is not a woman. If none of them are, then at least one isn’t!

Notice that these relationships are depicted in a slightly different way from the others on the Square of Opposition: there’s an arrow pointing toward the bottom. This is because the relationship is not symmetrical. If the proposition on top is true, then the one on the bottom must also be true; but the reverse is not the case. If an I proposition is true—some sailors are pirates—it doesn’t follow that the corresponding A proposition—that all sailors are pirates—is true. Likewise, the truth of an O proposition—some surfers are not priests—does not guarantee the truth of the corresponding E proposition—that no surfers are priests.

Truth, as it were, travels down the side of the Square. Falsehood does not: if the universal proposition is false, that doesn’t tell us anything about the truth or falsehood of the corresponding particular. You could have a false A proposition—all men are priests—with a true corresponding I—some men are priests. But you could also have a false A proposition—all cats are dogs—whose corresponding I—some cats are dogs—is also false. Likewise, you could have a false E proposition—no men are priests—with a true corresponding O—some men are not priests. But you could also have a false E proposition—no whales are mammals—whose corresponding O—some whales are not mammals—is also false.

Falsehood doesn’t travel down the side of the Square, but it does travel up. That is, if a particular proposition—I or O—is false, then its corresponding universal proposition—A or E, respectively—must also be false. Think about it in the abstract: if it’s false that some S are P, that means that there’s not even one S that’s also a P; well in that case, there’s no way *all* the Ss are Ps! False I, false A. Likewise on the negative side: if it’s false that some S are not P, that means you won’t find even one S that’s not a P, which is to say *all* the Ss are Ps; in that case, it’s false that no S are P (A and E are contraries). False O, false E.

## Inferences

Given information about the truth or falsity of a categorical proposition, we can use the relationships summed up in the Square of Opposition to make inferences about the truth-values of the other three types of categorical proposition.

Here’s what I mean. Suppose a universal affirmative proposition—an A proposition—is true. What are the truth-values of the corresponding E, I, and O propositions? (By “corresponding”, I mean propositions with the same subject and predicate classes.) The Square can help us answer these questions. First of all, A is in the opposite corner from O—they’re contradictory. That means A and O have to have opposite truth-values. Well, if A is true, as we’re supposing, then the corresponding O proposition has to be false. Also, A and E are contraries. That means that they can’t both be true. Well, we’re supposing that the A is true, so then the corresponding E must be false. What about the I proposition? Three ways

1. And the universal propositions are called *superalterns*.

to attack this one, and they all agree that the I must be true: (1) I is the subaltern of A, so if A is true, then I must be true as well; (2) I is the contradictory of E, and we've already determined that E must be false, so I must be true; (3) I and O are subcontraries, meaning they can't both be false, and since we've already determined that O is false, it follows that I must be true.

Summing up: if an A proposition is true, the corresponding E is false, I is true, and O is false.

Let's try another one: suppose a universal negative, E proposition is true. What about the corresponding A, I, and O propositions? Well, again, A and E are contraries—can't both be true— so A must be false. I is the contradictory of E, so it must be false—the opposite of I's truth-value. And since O is subaltern to E, it must be true because E is.

If an E proposition is true, the corresponding A is false, I is false, and O is true.

Another. Suppose a particular affirmative, I proposition is true. What about the other three? Well, E is its contradictory, so it must be false. And if some S are P, that means some of them aren't— so the O is also true. And since A is the contradictory of O... WAIT JUST A MINUTE! Go back and read that again. Do you see what happened? "And if some S are P, that means some of them aren't..." No it doesn't! Remember, 'some' means 'there is at least one'. If some S are P, that just means at least one S is a P—and for all we know, *all* of them might be; and then again, maybe not. I and O are subcontraries: they can't both be false, they *could* both be true, and one could be true and the other false. Knowing that I is true tells us nothing about the truth-value of the corresponding O, or the corresponding A. That some are, meaning at least one is, leaves open the possibility that all of them are; but then again, maybe not. The fact is, based on the supposition that an I is true, we can only know the truth-value of the corresponding E for sure.

If an I proposition is true, then the corresponding E is false, and A and O are of unknown truth value.

## EXERCISES

1. Suppose an O proposition is true. What are the truth-values of the corresponding A, E, and I propositions, according to the Square of Opposition?
2. Suppose an A proposition is false. What are the truth-values of the corresponding E, I, and O propositions, according to the Square of Opposition?
3. Suppose an E proposition is false. What are the truth-values of the corresponding A, I, and O propositions, according to the Square of Opposition?
4. Suppose an I proposition is false. What are the truth-values of the corresponding A, E, and O propositions, according to the Square of Opposition?
5. Suppose an O proposition is false. What are the truth-values of the corresponding A, E, and I propositions, according to the Square of Opposition?

## IV. Operations on Categorical Sentences

We continue our exploration of the portion of natural language to which Aristotle's logic restricts itself—the standard form sentences expressing categorical propositions. To familiarize ourselves more intimately with these, we will look at how they respond when we perform various operations on them, when we manipulate them in various ways. We will examine three operations: conversion, obversion, and contraposition. Each of these alters the standard form sentences in some way. The question we will ask is whether the new sentence that results from the manipulation is equivalent to the original sentence; that is, does the new sentence express the same proposition as the original?

## Conversion

Performing conversion on a categorical sentence involves changing the order of the subject and predicate terms. The result of this operation is a new sentence, which is said to be the *converse* of the original sentence. Our question is: when does performing conversion produce an equivalent new sentence, a converse that expresses the same proposition as the converted original? We will look at all four types standard form sentence, answering the question for each.

Let's perform conversion on a sentence expressing a universal affirmative, A proposition and see what happens. 'All dogs are animals' is such a sentence. conversion switches the subject and predicate terms, so the converse sentence is 'All animals are dogs'. Does the converse express the same proposition as the original? Are they equivalent? Heck, no! The original sentence expresses the true proposition that all dogs are animals; the converse expresses the utterly false proposition that all animals are dogs. Converting an A sentence produces a new sentence that is not equivalent to the original.

This means that the effect on truth-value, in the abstract, of converting A sentences, is unpredictable. Sometimes, as with 'All dogs are animals', conversion will lead you from a truth to a falsehood. Other times, it may lead from truth to truth: 'All bachelors are unmarried men' and 'All unmarried men are bachelors' express different propositions, but both of them are true (because it so happens that, by definition, a bachelor is just an unmarried man). conversion of an A could also lead from falsehood to falsehood, as with the transition from 'All dogs are bats' to 'All bats are dogs'. And it could lead from falsehood to truth: just reverse the order of the first conversion we looked at, from 'All animals are dogs' to 'All dogs are animals'.

Again, the point here is that, because conversion of A sentences produces a converse that expresses a different proposition than the original, we cannot know what the effect of the conversion will be on truth-value.

How about conversion of sentences expressing universal negative, E propositions? 'No dogs are cats' is such a sentence. Its converse would then be 'No cats are dogs'. Are they equivalent? Yes, of course. Remember, an E proposition denies even partial inclusion; it makes the claim that the two classes involved don't have any members in common. It doesn't matter which of the two classes is listed first in the sentence expressing that proposition—you still get the assertion that the two classes are exclusive. This is true of E sentences generally: performing conversion on them always produces a new sentence that is equivalent to the original.

It is also true of sentences expressing particular affirmative, I propositions. 'Some sailors are pirates', after conversion, becomes 'Some pirates are sailors'. These express the same proposition: they make the claim that the two classes have at least one member in common—there is at least one thing that is both a sailor and a pirate. Again, it doesn't matter what order you put the class terms in; I sentences express the assertion that there's overlap between the two classes. An I sentence and its converse are always equivalent.

The same cannot be said of sentences expressing particular negative, O propositions. Consider 'Some men are not priests'. That expresses a true proposition. But its converse, 'Some priests are not men' expresses a different proposition; we know it's a different proposition because it's false. That is all we need to show that an operation does not produce equivalent sentences: one counterexample. As above with A sentences, this means that the effect on truth-value of converting O sentences is unpredictable. It can take us from truth to falsehood, as in this example, or from truth to truth, falsehood to falsehood, falsehood to truth. In the abstract, we cannot know the effect on truth of converting O sentences, since the converse expresses a different proposition from the original.

Summary for conversion: for E and I, converses are equivalent; for A and O, converses are not.

## Obversion

Before we talk about our next operation, obversion, we need to introduce a new concept: class complements. The *complement* of a class, call it S, is another class which contains all the things that are not members of S. So, for example, the complement of the class of trees is just all the things that aren't trees. The easiest way to name class complements is just to stick the prefix 'non' in front of the original class name. So the complement of trees is non-trees.

Be careful: it may be tempting, for example, to say that the complement of Republicans is Democrats. But that's not right. The complement of Republicans is a much bigger class, containing *all* the non-Republicans: not just Democrats, but Communists and Libertarians and Independents and Greens; oh, and a bunch of other things, too—like the planet Jupiter (not a Republican), my left pinkie toe, the Great Wall of China, etc., etc.

As a matter of notational convention, if we use a capital letter like S to refer to a class, we will denote the complement of that class as  $\sim S$ , which we'll read as "tilde-S."

Back to obversion. Here's how this operation works: first, you change the quality of the sentence (from affirmative to negative, or *vice versa*); then, you replace the predicate with its complement. The result of performing obversion on a sentence is called the *obverse* of the original.

It turns out that performing obversion on a sentence always produces a new sentence that's equivalent to it; a sentence and its obverse always express the same proposition. That means they share a truth-value: if a sentence is true, so is its obverse; if a sentence is false, its obverse is false, too. We can see that this is so by looking at the result of performing obversion on each of the four types of standard form sentences.

We'll start with A sentences. Consider 'All ducks are swimmers'. To perform obversion on this sentence, we first change its quality. This is a universal affirmative. Its quality is affirmative. So we change that to negative, keeping the quantity (universal) the same. Our new sentence is going to be a universal negative, E sentence—something of the form No S are P. Next, we replace the predicate with its complement. The predicate of the sentence is 'swimmers'. What's the complement of that class? All the things that aren't swimmers: non-swimmers. So the obverse of the original A sentence is this: 'No ducks are non-swimmers'.

Now, are these two sentences equivalent? Yes. 'All ducks are swimmers' expresses the universal affirmative proposition, asserting that the class of ducks is entirely contained in the class of swimmers. That is to say, any duck you find will also be in the swimmer class. Another way of putting it: you won't find any ducks who *aren't* in the class of swimmers. In other words, no ducks fail to be swimmers. Or: 'No ducks are non-swimmers'. The A sentence and its obverse are equivalent; they express the same proposition, make the same claim about the relationship between the class of ducks and the class of swimmers.

Let's try obversion on a universal negative, E sentence. 'No women are priests' is one. First, we change its quality from negative to affirmative: it becomes a universal affirmative, A sentence— something of the form All S are P. Next, we replace its predicate, 'priests', with its complement, 'non-priests'. The result: 'All women are non-priests'. Is that equivalent to the original? It tells us that all women are outside the class of priests. In other words, none of them are priests. That is, 'No women are priests'. Yes, both the original sentence and its obverse tell us that the classes of women and priests are exclusive.

Next, the particular affirmative—an I sentence like 'Some politicians are Democrats'. OK. First, change the quality—from affirmative to negative. Our obverse will be a particular negative, O sentence—something of the form Some S are not P. Now, replace 'Democrats' with 'nonDemocrats', stick it in the predicate slot, and we get 'Some politicians are not non-Democrats'. Well, that's not exactly grammatically elegant, but the meaning is clear: not being a non-Democrat is just being a Democrat. This says the same as the original, namely that some politicians are Democrats.

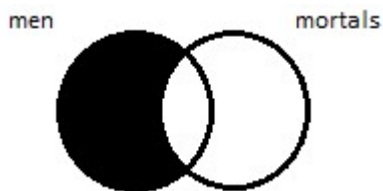
Finally, particular negative, O. We'll try 'Some plants are not flowers'. Changing from negative to affirmative means our obverse will be an I—Some S are P. We replace 'flowers' with 'nonflowers' and get 'Some plants are non-flowers'. We went from 'Some plants are not flowers' to 'Some plants are non-flowers'. Obviously, those are equivalent.

Summary for obversion: obverses are equivalent for A, E, I, and O.

## Contraposition

Our last operation is contraposition. Unlike obversion, and like conversion, it doesn't involve changing the type (A, E, I, O) of the sentence we're operating on. Rather, again, like conversion, we just manipulate the subject and predicate. Here's how: replace the subject with the complement of the predicate; and replace the predicate with the complement of the subject. The result of performing contraposition on a sentence is called its *contrapositive*.

Let's perform contraposition on an A sentence: 'All men are mortals'. To form its contrapositive, we put the complement of the predicate—non-mortals—into subject position and the complement of the subject—non-men—into predicate position: 'All non-mortals are non-men'. The question, as always: are these sentences equivalent? This one's a bit hard to see. Let's use Venn diagrams to help us think it through. First, we know what the diagram for 'All men are mortals' looks like; that sentence claims that there's no such thing as a man who's not a mortal, so we blot out the portion of the 'men' circle that's not inside the 'mortals' circle:

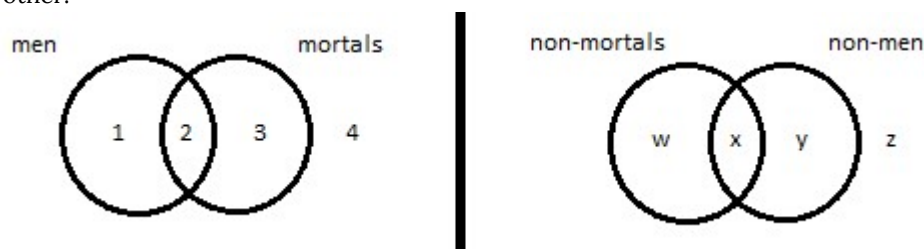


Next, let's think through how we would diagram 'All non-mortals are non-men'. If we change our circles to 'non-men' and 'non-mortals', respectively, it's easy; when you're diagramming an A proposition, you just blot out the part of the left-hand (subject) circle that doesn't overlap with the right-hand (predicate) circle. There's no such thing as non-men who aren't non-mortals:



But how do we compare this diagram with the one for 'All men are mortals' to see if they express the same proposition? We need to know that the two would give us the same picture if the circles were labeled the same.

Let's compare the unshaded diagrams where the circles are 'men' and 'mortals', on the one hand, and 'non-men' and 'non-mortals' on the other:



When we depict 'All men are mortals', we blot out region 1 of the left-hand diagram. When we depict its contrapositive, 'All non-mortals are non-men', we blot out region w of the right-hand diagram. We want to know whether these two sentences are equivalent. They are, provided that blotting out region 1 and blotting out region w amount to the same thing. Do they? That is, do regions 1 and w contain the same objects?

Let's think this through, starting with region z. What's in there? Those are the things that are outside both the non-mortal and non-men circles; that is, they're *not* non-mortals and they're *not* non-men. So they're mortals and men, right? Things that are both mortals and men: on the lefthand diagram, that's the overlap between the circles. Region z and region 2 contain the same things.

How about region y? Those things are non-men, but they're outside the non-mortals circle, making them mortals. Mortals who aren't men: they live in region 3 in the left-hand diagram. Regions y and 3 contain the same things. Region x has things that are both non-men and non-mortals; that is, they're outside both the mortal and men circles on the left. Regions x and 4 contain the same things.

And region w? Outside the non-men circle, so they're men. Inside the non-mortals circle, so they're not mortals. Men that aren't mortals: that's region 1 on the left. Regions w and 1 contain the same things. And that means that blotting

out region w and blotting out region 1 amount to the same thing; both are ways of ruling out the existence of the same group of objects, the men who aren't mortals—or, as it turns out, the non-mortals who aren't non-men. Same thing.

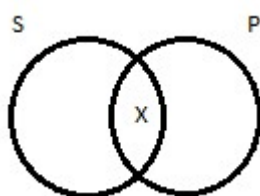
Picking the main thread back up, what all this shows is that when we perform contraposition on universal affirmative, A sentences, we end up with new sentences that express the same proposition. An A sentence and its contrapositive are equivalent. We still have to ask the same question about E, I, and O sentences.

Consider a universal negative (E): 'No sky-divers are cowards'. This is surely true; it takes bravery to jump out of a plane (I wouldn't do it). To get the contrapositive, we replace the subject, skydivers, with the complement of the predicate, non-cowards; and we replace the predicate, cowards, with the complement of the subject, non-sky-divers. The result is 'No non-cowards are non-skydivers'. That's false. You know who was a non-coward? Martin Luther King, Jr. The Reverend King was a courageous advocate for racial equality up to the very last day of his life. But, not a sky-diver. The contrapositive claims there's no such thing as a non-coward who doesn't sky-dive. But that ain't so: MLK is a counterexample. In general, when you perform contraposition on an E sentence, you end up with a new sentence that expresses a different proposition. And as was the case with A and O sentences being converted, this has unpredictable effects on truth-value. You may move from truth to falsehood, as in this case, or from truth to truth, falsehood to falsehood, falsehood to truth. Contraposition changes the proposition expressed by E sentences, so you can't know.

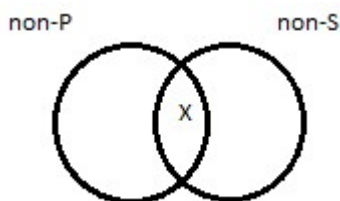
Next, consider particular negative (O) sentences. These are pretty easy. 'Some men are not priests' is a good go-to example. Performing contraposition, we get 'Some non-priests are not non-men'. Things that are not non-men—those are just men. So the claim being made by the contrapositive is that some non-priests are men. That is, there's at least one thing that's both a non-priest and a man; or, there's at least one man who's not a priest. I know a way to say that: 'Some men are not priests'. The O sentence and its contrapositive make the same claim. Contraposition performed on particular negatives gives you a new sentence that is equivalent to the original.

Finally, particular affirmatives—I sentences. 'Some men are priests' is true. So is its contrapositive: 'Some non-priests are non-men' (there's at least one: my mom is not a man, nor was she ever a priest). So contraposition performed on an I works? That is, it gives you an equivalent sentence? Not necessarily. The two sentences might both be true, but they could be expressing two different true propositions. As a matter of fact, they are. When you contrapose an I sentence, the result is a new sentence that is not equivalent. To see why, we'll return to Venn diagrams.

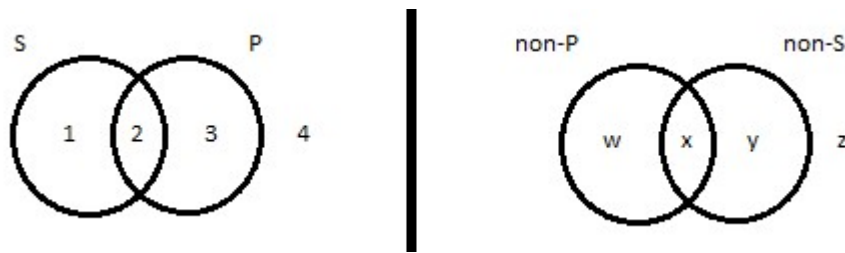
Generically speaking, an I proposition's diagram has an X in the area of overlap between the two circles. For a sentence of the form Some S are P, we would draw this:



There is at least one thing (the X) that is both S and P. For the contrapositive, we draw this:



There is at least one thing that is both non-P and non-S. The question is, does drawing an X in those two regions of overlap amount to the same thing? Let's put the diagrams side-by-side, without the Xs, but with numbers and letters for the different regions:



We went through this above when we were discussing the effects of contraposition on A propositions. Regions 1 and w contain the same things, as do regions 3 and y. But regions 2 and 4 don't line up with regions x and z, respectively. Rather, they're reversed: region 2 has the same objects as region z, and region 4 has the same objects as region x.

When we draw the picture of the straight-up I sentence, we put an X in region 2; when we draw the picture of its contrapositive, we put an X in region x. But region 2 and region x aren't the same. So the I sentence and its contrapositive, in general, are not equivalent. Performing contraposition on an I sentence changes the proposition expressed, with unpredictable effects on truth-value.

We can prove it with a concrete example. Let our starting I sentence be 'Some Catholics are nonPopes'. That's certainly true (again, my mom: Catholic, but not Pope). The contrapositive would be 'Some Popes are non-Catholics' (the complement of non-Popes is just Popes). But that's false. Being Catholic is a prerequisite for the Papacy. An I sentence and its contrapositive make different claims.

## EXERCISES

- Perform conversion on the following and write down the converse. Is it equivalent to the original sentence?
  - Some surfers are not priests.
  - All Canadians are bodybuilders.
  - No Mexicans are fishermen.
  - Some Nazis are florists.
- Perform obversion on the following and write down the obverse. Is it equivalent to the original sentence?
  - No people are lizards.
  - Some politicians are criminals.
  - Some birds are not animals.
  - All Democrats are samurais.
- Perform contraposition on the following and write down the contrapositive. Is it equivalent to the original sentence?
  - All Philistines are Syrians.
  - No Africans are Europeans.
  - Some Americans are Irishmen.
  - Some Swiss are not Catholics.

## *Inferences*

Earlier, we discussed how we could make inferences about the truth-values of categoricals using the information encoded in the Square of Opposition. For example, given the supposition that an A sentence expresses a true proposition, we can infer that the corresponding E sentence expresses a falsehood (since A and E are contraries, which can't both be true), that the corresponding I sentence expresses a truth (since I is the subaltern of A, which means A's truth guarantees that of I), and that the corresponding O sentence expresses a falsehood (since A and O are contradictories, which must have opposite truth-values).



The key word in that paragraph is ‘corresponding’. The Square of Opposition tells us about the relationships among categoricals that *correspond*—which means they have the same subjects and predicates. If ‘All S are P’ is true, then ‘No S are P’ must be false, per the Square, since these two sentences have the same subject (S) and predicate (P). The square *cannot* license such inferences when the subjects and predicates do not correspond. The supposition that ‘All S are P’ is true tells me nothing at all about the truth-value of ‘Some A are B’; the subjects and predicates are different; we’re dealing with two different classes.

There are occasions, however, when subjects and predicates do not correspond, but we can nevertheless make inferences about the truth-values of categoricals based on information about others. In such cases, we need to combine our knowledge of the relationships depicted in the Square of Opposition with our recently acquired knowledge about the circumstances in which conversion, obversion, and contraposition provide us with equivalent sentences.

Here is a simple example. Suppose that a sentence of the form ‘No S are P’ expresses a truth (never mind what ‘S’ and ‘P’ stand for; we’re thinking in the abstract here). Given that information, what can we say about a sentence of the form ‘Some P are S’? Well, the first is an E and the second is an I. According to the Square of Opposition, E and I are a contradictory pair, so they must have opposite truth-values. But remember, the relationships in the Square only hold for *corresponding* sentences. ‘No S are P’ and ‘Some P are S’ do not correspond; their subject and predicate class terms are in different spots. The Square tells us that the I sentence corresponding to ‘No S are P’—namely, ‘Some S are P’—must be the opposite truth-value. We’ve presumed that the E sentence is true, so ‘Some S are P’ expresses a falsehood, according to the Square. But we wanted to know the truth-value of ‘Some P are S’, the sentence with the subject and predicate terms switched. Well, switched subject and predicate terms—that’s just the *converse* of ‘Some S are P’. And we know from our investigations that performing conversion on an I sentence always gives you another I sentence that’s equivalent to the first; that is, it expresses the same proposition, so it’s true or false in all the same circumstances as the original. That means ‘Some P are S’ must express a falsehood, just like its converse.

Here’s how to think about the inference we just made. We were given the fact that ‘No S are P’ is true. We wanted to know the truth-value of ‘Some P are S’. We can’t compare these two directly using the Square of Opposition because they don’t correspond: different subject and predicate. But, we know that the converse of the our target sentence—‘Some S are P’—*does* correspond, so according to the Square, it must be false (since it’s contradictory to ‘No S are P’). And, since conversion on I sentences yields equivalent results, ‘Some P are S’ has the same truth-value as ‘Some S are P’, so our target sentence must also be false.<sup>2</sup>

This is the general pattern for these sorts of multi-step inferences. You’re given information about a particular categorical claim’s truth-value, then asked to evaluate some other claim for truth or falsity. They may not correspond, so the first stage of your deliberations involves getting them to correspond—making the subject and predicate terms line up. You do this by performing conversion, obversion, and contraposition as needed, but only when those operations produce equivalent results: you only use conversion on E and I sentences; you only use contraposition on A and O sentences; and since obversion always yields an equivalent sentence, you can use it whenever you want. Then, once you’ve achieved correspondence, you can consult the Square of Opposition and complete the inference.

Another example can help illustrate the method. Suppose we’re told that some sentence ‘All S are P’ is true. What about the sentence ‘No ~ S are ~ P’? (Remember, when we put the tildes in front of the letters, we’re referring to the complements of these classes.)

First, we notice that the subject and predicate terms don’t correspond. The A sentence has ‘S’ in subject position and ‘P’ in predicate position, while the target E sentence has ~ S and ~ P in those slots. We can see this misalignment clearly

2. We’re getting a little sloppy here. Technically, it’s propositions, not sentences, that are true or false. Further complication: we’re not even talking about actual sentences here, but generic sentence-patterns, with placeholder letters ‘S’ and ‘P’ standing in for actual class terms. Can those sorts of things be true or false? Ugh. We could precise things by saying, “would be true/false were it instantiated with a propositional constant . . .” but this is quite the mouthful.. Let’s just agree not to be fussy and not to worry about it. We all understand what’s going on.

(and also set ourselves up more easily to think through the remaining steps in the inference) if we write the sentences out, one above the other (noting in brackets what we know about their truth-values):

All S are P [T]  
No ~ S are ~ P [?]

Focusing only on subject and predicate terms, we see that the bottom ones have tildes, the top ones don't. We need to get them into correspondence. How? Well, it occurs to me that we have an operation that allows us to add or remove tildes, two at a time: contraposition. When we perform that operation, we replace the subject with the complement of the predicate (adding or removing one tilde) and we replace the predicate with the complement of the subject (adding or removing another). Now, contraposition produces equivalent sentences for A and O, but not E and I. So I can only perform it on the top sentence, 'All S are P'. Doing so, I produce a contrapositive which expresses the same proposition, and so must also be true. We can write it down like this:

<hr/>	
All S are P	[T]
	[T]
All ~ P are ~ S	
No ~ S are ~ P	[?]
<hr/>	

The sentence we just wrote down still doesn't align with the target sentence at the bottom, but it's closer: they both have tildes in front of 'S' and 'P'. Now the problem is that the '~ S' and '~ P' are in the wrong order: subject and predicate positions, respectively, in the target sentence, but the reverse in the sentence we just wrote down. We have an operation to fix that! It's called conversion: to perform it, you switch the order of subject and predicate terms. The thing is, it only works— that is, gives you an equivalent result—on E and I sentences. I can't perform conversion on the A sentence 'All ~ P are ~ S' that I just wrote down at the top. But, I can perform it on the target E sentence at the bottom:

<hr/>	
All S are P	[T]
	[T]
All ~ P are ~ S	
No ~ P are ~ S	[?]
No ~ S are ~ P	[?]
<hr/>	

I did conversion, as it were, from the bottom up. Those last two E sentences are converses of one another, so they express the same proposition and will have the same truth value. If I can figure out the truth-value of 'No ~ P are ~ S', then I can figure out the truth-value of my target sentence on the bottom; it'll be the same. And look! I'm finally in a position to do that. The two sentences in the middle, 'All ~ P are ~ S' and 'No ~ P are ~ S', correspond; they have the same subject and predicate. That means I can consult the Square of Opposition. I have an A sentence that's true. What about the corresponding E sentence? They're contraries, so it must be false:

<hr/>	
All S are P	[T]
	[T]
All ~ P are ~ S	
No ~ P are ~ S	[F]
No ~ S are ~ P	[?]
<hr/>	

And since the target sentence at the bottom expresses the same proposition as the one directly above it, that final question mark can also be replaced by an 'F'. Inference made, problem solved.

Again, this is the general pattern for making these kinds of inferences: achieve correspondence by using the three operations, then use the information encoded in the Square of Opposition.

This works most of the time, but not always. Suppose you're told that 'All S are P' is true, and asked to infer the truth-value of 'No P are ~ S'. We can again write them out one above the other and take a look:

All S are P [T]  
No P are ~ S [?]

'S' and 'P' are in the wrong order, plus 'S' has a tilde in front of it on the bottom but not on the top. The first thing that occurs to me to do is to get rid of that tilde. We have an operation for adding or removing one tilde at a time: obversion. I'm going to perform it on the bottom sentence. First, I change the quality: the universal negative (E) original becomes a universal affirmative (A). Then I replace the predicate with its complement: I replace '~ S' with just plain 'S'. This is the result:

All S are P [T]  
All P are S [?]  
No P are ~ S [?]

We don't have correspondence yet, but we're closer with that tilde out of the way. What next? Well, now the problem is just that 'S' and 'P' are in the wrong order. There's an operation for that: conversion. But—and here's the rub—we can only use conversion on E and I sentences. Now that I did obversion on the target at the bottom, the two sentences I'm left comparing are both As. I can't use conversion on an A: the result won't be equivalent.

At this point, the sensible thing to do would be to try other operations: maybe the right combination of obversion, contraposition, and possibly, eventually, on a different kind of sentence, conversion, will allow us to achieve correspondence. When making these kinds of inferences, you often have to try a variety of things before you get there. But I'm here to tell you, try what you might in this example, as many conversions, obversions, and contrapositions as you want, in whatever order: you'll never achieve correspondence. It's impossible.

So what does that mean? It means that, given the fact that 'All S are P' is true, you cannot make any inference about the truth-value of 'No P are ~ S'. The answer to the problem is: "I don't know." Remember, this kind of thing can happen; sometimes we can't make inferences about one categorical based on information about another. When we know that an I is true, for example, we can't say what the truth-value of the corresponding O is; it could go either way.

That's kind of unsatisfying, though. I'm telling you that if you can't achieve correspondence—if it's impossible—that you can't make an inference. But how do you *know* that you can't achieve correspondence? Maybe, as you were laboring over the problem, you just didn't stumble on the right combination of operations in the right order. How do we know for sure that an inference can't be made? As a matter of fact, the one step that we took in this problem puts us in a position to know just that. Compare 'All S are P' with the obverse of the target sentence, 'All P are S'. What's the relationship between those? One is the converse of the other. We're given a true A sentence, and asked to make an inference about the truth-value of a sentence equivalent to its converse. But performing conversion on an A, as we established at length above, gives you a new sentence that expresses a different proposition. And this has unpredictable effects on truth-value: sometimes one goes from truth to falsity; other times from truth to truth, and so on. In this case, we *know* that we can't know the truth-value of the target sentence, because it's equivalent to the result of perform conversion on a universal affirmative, and the effects of that operation on truth-value are unpredictable.

In general, you can know that the answer to one of these problems is "I don't know" if you can use the operations to get into a position where you're comparing a sentence with its converse or contrapositive when those operations don't work for the types of sentences you have. We saw this for an A and its converse. Similarly, if you have an E sentence of known truth-value, and your target sentence is equivalent to its contrapositive, you know the answer is "I don't know," because contraposition performed on E sentences has unpredictable results on truth-value. Same goes for I and conversion, O and contraposition.

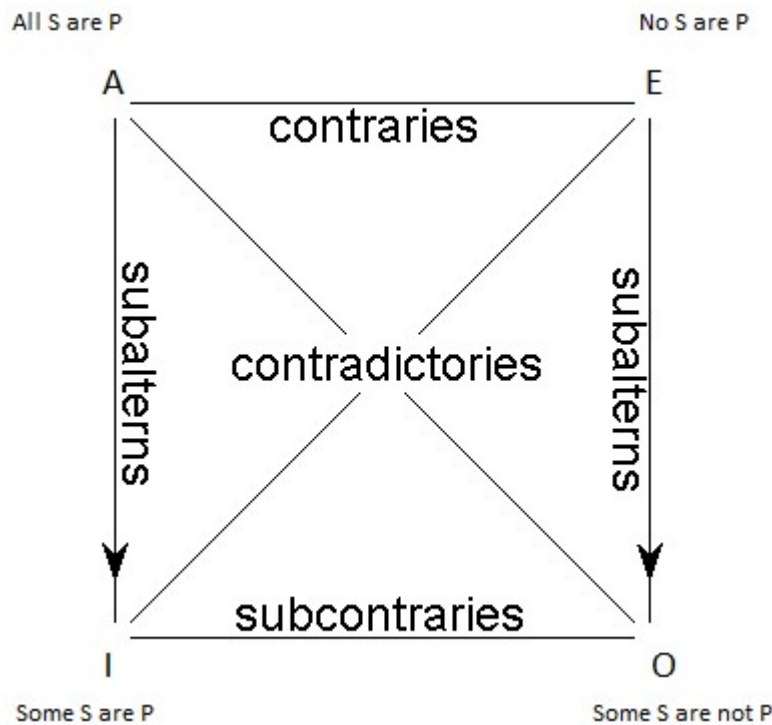
## EXERCISES

1. Suppose 'All S are P' is true. Determine the truth-values of the following (if possible).
  - (a) No S are  $\sim P$
  - (b) All  $\sim S$  are  $\sim P$
  - (c) No  $\sim P$  are S
  - (d) Some  $\sim P$  are S
  - (e) Some  $\sim S$  are not  $\sim P$
2. Suppose 'No S are P' is true. Determine the truth-values of the following (if possible).
  - (a) Some  $\sim P$  are not  $\sim S$
  - (b) All  $\sim S$  are  $\sim P$
  - (c) No  $\sim S$  are  $\sim P$
  - (d) Some  $\sim P$  are S
  - (e) All  $\sim P$  are  $\sim S$
3. Suppose 'Some S are P' is true. Determine the truth-values of the following (if possible).
  - (a) All S are  $\sim P$
  - (b) Some S are not  $\sim P$
  - (c) No P are S
  - (d) Some P are  $\sim S$
  - (e) No S are  $\sim P$
4. Suppose 'Some S are not P' is true. Determine the truth-values of the following (if possible).
  - (a) No S are  $\sim P$
  - (b) Some S are  $\sim P$
  - (c) No  $\sim S$  are P
  - (d) No  $\sim P$  are S
  - (e) Some P are S

### *V. Problems with the Square of Opposition*

The Square of Opposition is an extremely useful tool: it neatly summarizes, in graphical form, everything we know about the relationships among the four types of categorical proposition.

Except, actually, we don't know those things. I'm sorry, but when I first presented the Square of Opposition and made the case for the various relationships it depicts, I was leading you down the proverbial primrose path. What appeared easy is in fact not so simple as it seems. Some of the relationships in the Square break down under certain circumstances and force us to do some hard thinking about how to proceed. It's time to explore the "steep and thorny way" that opens before us when we dig a bit deeper into problems that can arise for the Square of Opposition.



## Existential Import

To explain what these problems are, we need the concept of *existential import* (E.I. for short). E.I. is a property that propositions may or may not have. A proposition has existential import when its truth implies the existence of something. Because of what we decided to mean when we use the word ‘some’—namely, ‘there is at least one’—the particular propositions I and O clearly have E.I. For ‘Some sailors are not pirates’ to be true, there has to exist at least one sailor who is not a pirate. Again, that’s just a consequence of what we mean by ‘some’.

In addition, given the relationships that are said to hold by the Square of Opposition, the universal propositions A and E also have existential import. This is because the particular propositions are subalterns. The truth of a universal proposition implies the truth of a particular one: if an A is true, then the corresponding I must be; if an E is true, then the corresponding O must be. So since the truth of universals implies the truth of particulars, and particulars have E.I., then universals imply the existence of something as well: they have existential import, too.

## Problems for the Square

OK, all four of the categorical propositions have existential import. What’s the big deal? Well, this fact leads to problems. Consider the proposition that all C.H.U.D.s are Idahoans; also, consider the proposition that some C.H.U.D.s are not Idahoans. Both of these propositions are false. That’s because both of them imply the existence of things—namely, C.H.U.D.s—that don’t exist. (‘C.H.U.D.’ stands for ‘Cannibalistic Humanoid Underground Dweller’. They’re the titular scary monsters of a silly horror movie from the ’80s. They’re not real.) ‘Some C.H.U.D.s are not Idahoans’ claims that there exists at least one C.H.U.D. who’s not an Idahoan; but that’s not the case, since there are no C.H.U.D.s. ‘All C.H.U.D.s are Idahoans’ is also false: if it were true, its subaltern ‘Some C.H.U.D.s are Idahoans’ would have to be true; but it can’t be, because it claims that there’s such a thing as a C.H.U.D. (who’s an Idahoan).

Bottom line: A and O propositions about C.H.U.D.s both turn out false. This is a problem for the Square of Opposition because A and O are supposed to be a contradictory pair; they’re supposed to have opposite truth-values.

It gets worse. Any time your subject class is empty—that is, like ‘C.H.U.D.s’, it doesn’t have any members—all four of the categorical propositions turn out false. This is because, as we saw, all four have existential import. But if E and I are both false, that’s a problem: they’re supposed to be contradictories. If I and O are both false, that’s a problem: they’re supposed to be subcontraries. When we talk about empty subject classes, the relationships depicted in the Square cease to hold.

### **Solution?**

So the problems are caused by empty classes. We can fix that. We’re building our own logic from the ground up here. Step one in that process is to tame natural language. The fact that natural language contains terms that don’t refer to anything real seems to be one of the ways in which it is unruly, in need of being tamed. Why not simply restrict ourselves to class terms that actually refer to things, rule out empty classes? Then the Square is saved.

While tempting, this solution goes too far. The fact is, we make categorical claims using empty (or at least possibly empty) class terms all the time. If we ruled these out, our ability to evaluate arguments containing such claims would be lost, and our logic would be impoverished.

One field in which logic is indispensable is mathematics. Mathematicians need precise language to prove interesting claims. But some of the most interesting claims in mathematics involve empty classes. For instance, in number theory, one can prove that there is no largest prime number—they go on forever. In other words, the term ‘largest prime number’ refers to an empty class. If our logic ruled out empty class terms, mathematicians couldn’t use it. But mathematicians are some of our best customers!

Also, physicists. Before its existence was confirmed in 2013, they made various claims about a fundamental particle called the Higgs boson. “Higgs bosons have zero spin,” they might say, making a universal affirmative claim about these particles. But before 2013, they didn’t even know if such particles existed. IF they existed, they would have zero spin (and a certain mass, etc.); the equations predicted as much. But those equations were based on assumptions that may not have been true, and so there may not have been any such particle. Nevertheless, it was completely appropriate to make claims about it, despite the fact that ‘Higgs boson’ might be an empty term.

We make universal claims in everyday life that don’t commit us to the existence of things. Consider the possible admonition of a particularly harsh military leader: Deserters will be shot. This is a universal affirmative claim. But it doesn’t commit to the existence of deserters; in fact, it’s very purpose is to ensure that the class remains empty!

So, empty classes have their uses, and we don’t want to commit ourselves to the existence of things every time we assert a universal claim. Ruling out empty classes from our logic goes too far to save the Square of Opposition. We need an alternative solution to our problems.

## **Boolean Solution**

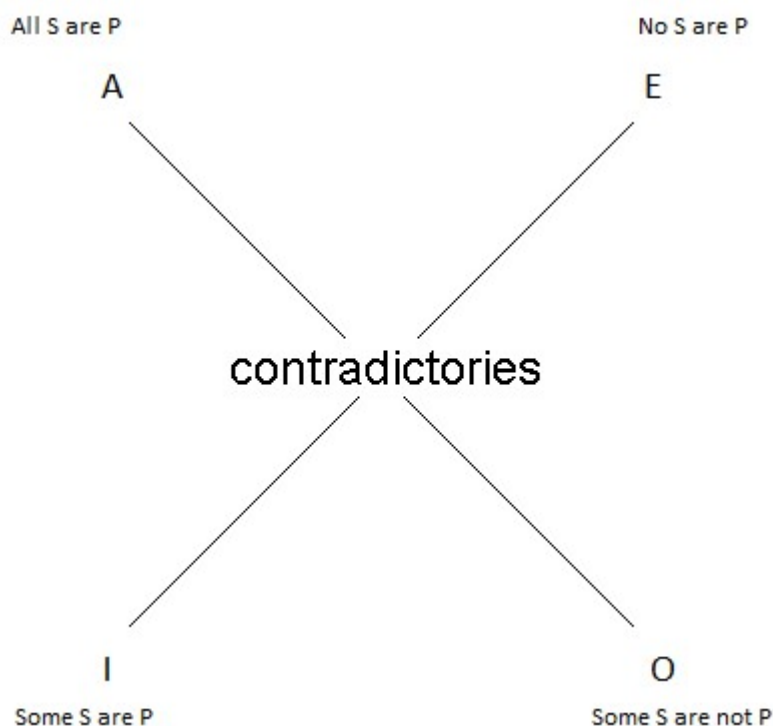
Advocated by the English logician George Boole in the 19<sup>th</sup> Century, our solution to the problems raised will be to abandon the assumption that universal propositions (A and E) have existential import, allow empty classes, and accept the consequences. Those consequences, alas, are quite dire for the traditional Square of Opposition. Many of the relationships it depicts do not hold when subject classes are empty.

First, the particular propositions (I and O) are no longer subcontraries. Since they start with the word ‘some’, they have existential import. When their subject classes are empty, as is now allowed, they both turn out false. Subcontraries can’t both be false, but I and O can both be false when we allow empty classes.

Next, the particular propositions are no longer subalterns of their corresponding universals (A and E). As we said, the universals no longer have existential import—they no longer imply the existence of anything—and so their truth cannot imply the truths of particular propositions, which do continue to have E.I.

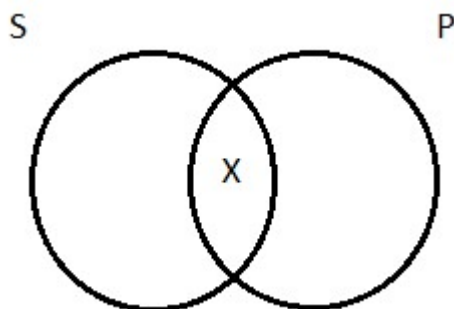
The only two relationships left on the Square now are contradictoriness—between A and O, E and I—and contrariety between the two universals. And these are in conflict when we have empty subject classes. In such cases, both I and O are false, as we’ve said. It follows that their contradictories, A and E, must be true. But, A and E are supposed to be a contrary pair; they can’t both be true. So we can’t keep both contrariety and contradictoriness; one must go. We will keep contradictoriness. To do otherwise would be to abandon the plain meanings of the words we’re using.

There's a reason I introduced this relationship first: it's the easiest to understand. If you want to contradict my universal affirmative claim that all sailors are pirates, you claim that some of them aren't; A and O are clearly contradictory. As are E and I: if you want to contradict my claim that no surfers are priests, you show me one who is. So we eliminate contrariety: it is possible, in cases where the subject class is empty, for both A and E propositions to be true. What we're left with after making these revisions is no longer a square, but an X. All that remains is contradictoriness:



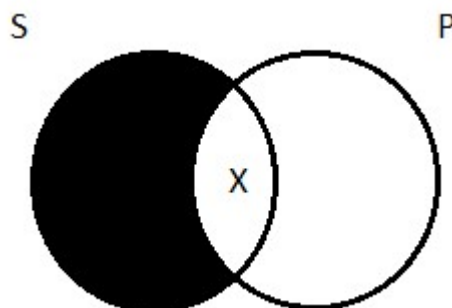
And our solution is not without awkwardness. In cases where the subject class is empty, both particular propositions (I and O) are false; their universal contradictories (E and A), then, are true in those circumstances. This is strange. Both of these sentences express truths: 'All C.H.U.D.s are Idahoans' and 'No C.H.U.D.s are Idahoans'. That's a tough pill to swallow, but swallow it we must, given the considerations above. We can make it a bit easier to swallow if we say that they're true, but vacuously or trivially. That is, they're true, but not in a way that tells you anything about how things actually are in the world (the world is, after all and thankfully, C.H.U.D.-free).

That we would end up choosing this interpretation of the categoricals, rather than the one under which universal propositions had existential import, was foreshadowed earlier, when we first introduced the four types of categorical proposition and talked about how to diagram them. We chose diagrams for A and E that did not imply the existence of anything. Recall that our way of indicating existence in Venn diagrams is to draw an X. So for a particular affirmative—some surfers are priests, say—we drew this picture (with the X being the one surfing priest we're committed to the existence of):

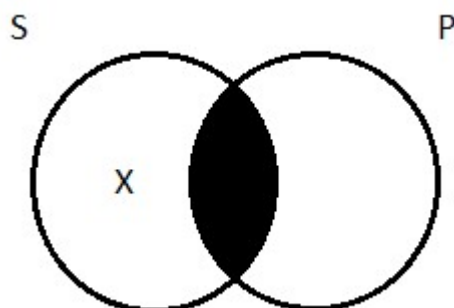


The diagrams for the universals (A and E), though, had no Xs in them, only shading; they don't commit us to the existence of anything. If we were going to maintain the existential import of A and E, we would've drawn different diagrams. For

the universal affirmative—all logicians are Germans, say—we'd shade out the portion of the left-hand circle that doesn't overlap the right, to indicate that there's no such thing as a logician who's not a German. But we would also put an X in the middle region, to indicate that there is at least one logician who is (existential import):



And for the universal negative—no women are priests, say—we would shade out the middle region, to indicate there's nothing that's both a woman and a priest. But we would also put an X in the left-hand circle, to indicate that there's at least one woman who's not a priest:



This interpretation of the universal propositions, according to which they have existential import, is often called the “Aristotelian” interpretation (as opposed to our “Boolean” interpretation, according to which they do not).<sup>3</sup> Which interpretation one adopts makes a difference. There are some arguments that the two interpretations evaluate differently: on the Aristotelian view, they are valid, but on the Boolean view, they are not. We will stick to the Boolean interpretation of the universals, according to which they do not have existential import.

## VI. Categorical Syllogisms

As we've said, Aristotelian Logic limits itself to evaluating arguments all of whose propositions— premises and conclusion—are categorical. There is a further restriction: Aristotelian Logic only evaluates categorical syllogisms. These are a special kind of argument, meeting the following conditions:

A *categorical syllogism* is a deductive argument consisting of three categorical propositions (two premises and a conclusion); collectively, these three propositions feature exactly three classes; each of the three classes occurs in exactly two of the propositions.

3. It is not clear, however, that it is correct to attribute this view to Aristotle. While he clearly did believe that universal affirmative (A) propositions had existential import, it's not clear that he thought the same about universal negatives. His rendering of the particular negative (O) was 'Not all S are P', which could be (trivially, vacuously) true when S is empty. In that case, O's being the subaltern of E does not force us to attribute Existential Import to the latter. For discussion, see Parsons, Terence, "The Traditional Square of Opposition", The Stanford Encyclopedia of Philosophy (Summer 2015 Edition), Edward N. Zalta (ed.), URL = <<http://plato.stanford.edu/archives/sum2015/entries/square/>>.



That's a mouthful, but an example will make it clear. Here is a (silly) categorical syllogism:

All chipmunks are athletes.  
Some athletes are golfers.  
/∴ Some chipmunks are  
golfers.

This argument meets the conditions in the definition: it has three propositions; there are exactly three classes involved (chipmunks, athletes, and golfers); and each of the three classes occurs in exactly two of the propositions (check it and see).

There is some special terminology for the class terms and premises in categorical syllogisms. Each of the three class terms has a special designation. The terms in the conclusion determine the designations. The so-called *major term* is the term that appears in predicate position in the conclusion; in our silly example, that's 'golfers'. The *minor term* is the term that appears in subject position in the conclusion; in our example, that's 'chipmunks'. The *middle term* is the other one, the one that appears in each of the premises; in our example, it's 'athletes'.

The premises have special designations as well. The *major premise* is the one that has the major term in it; in our example, that's 'Some athletes are golfers'. The *minor premise* is the other one, the one featuring the minor term; in our example, it's 'All chipmunks are athletes'.

Final restriction: categorical syllogisms must be written in standard form. This means listing the premises in the correct order, with the major premise first and the minor premise second. If you look at our silly example, you'll note that it's not in standard form. To fix it, we need to reverse the order of the premises:

Some athletes are golfers.  
All chipmunks are athletes.  
/∴ Some chipmunks are golfers.

An old concern may arise again at this point: in restricting itself to such a limited class of arguments, doesn't Aristotelian Logic run the risk of not being able to evaluate lots of real-life arguments that we care about? The response to this concern remains the same: while most (almost all) real-life arguments are not presented as standard form categorical syllogisms, a surprising number of them can be translated into that form. Arguments with more than two premises, for example, can be rewritten as chains of two-premise sub-arguments. As was the case when we raised this concern earlier, we will set aside the messy details of exactly how this is accomplished in particular cases.

## Logical Form

As we said at the outset of our exploration of deductive logic, there are three things such a logic must do: (1) tame natural language; (2) precisely define logical form; and (3) develop a way to test logical forms for validity. Until now, we've been concerned with the first step. It's (finally) time to proceed to the second and third.

The logical form of a categorical syllogism is determined by two features of the argument: its *mood* and its *figure*. First, mood. The mood of a syllogism is determined by the types of categorical propositions contained in the argument, and the order in which they occur. To determine the mood, put the argument into standard form, and then simply list the types of categoricals (A, E, I, O) featured in the order they occur. Let's do this with our silly example:

Some athletes are golfers.  
All chipmunks are athletes.  
/∴ Some chipmunks are golfers.

From top to bottom, we have an I, an A, and an I. So the mood of our argument is IAI. It's that easy. It turns out that there are 64 possible moods—64 ways of combining A, E, I, and O into unique three-letter combinations, from AAA to OOO and everything in between.

The other aspect of logical form is the argument's *figure*. The figure of a categorical syllogism is determined by the arrangement of its terms. Given the restrictions of our definition, there are four different possibilities for standard form

syllogisms. We will list them schematically, using these conventions: let 'S' stand for the minor term, 'P' stand for the major term, and 'M' stand for the middle term. Here are the four figures:

(1) MP	(2) PM	(3) MP	(4) PM
SM	SM	MS	MS
∴ SP	∴ SP	∴ SP	∴ SP

Again, the only thing that determines figure is the arrangement of terms—whether they appear in subject or predicate position in their premises. In our schemata, that the letter is listed first indicates that the term appears in subject position; that it appears second indicates that it's in predicate position. So, in the first figure, in the major premise (the first one), the middle term (M) is in subject position and the major term (P) is in predicate position. Notice that for all four figures, the subject and predicate of the conclusion remains the same: this is because, by definition, the minor term (S) is the subject of the conclusion and the major term (P) its predicate.

Returning to our silly example, we can determine its figure:

Some athletes are golfers.  
 All chipmunks are athletes.  
 ∴ Some chipmunks are golfers.

Perhaps the easiest thing to do is focus on the middle term, the one that appears in each of the premises—in this case, 'athletes'. It occurs in subject position in the major premise, then predicate position in the minor premise. Scanning the four figures, I just look for the one that has 'M' listed in first position on the top, then second position in the middle. That's the first figure.

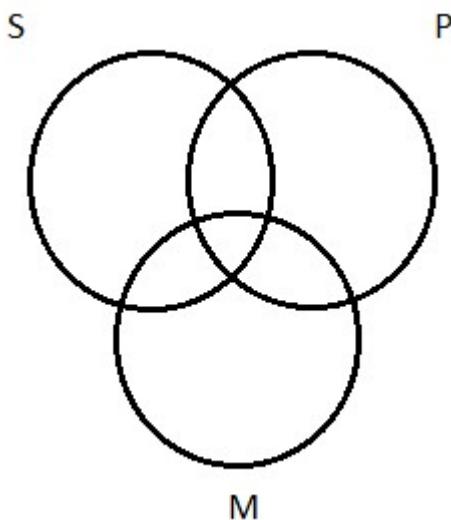
So the mood of our sample argument is IAI, and it's in the first figure. Logical form is just the mood and figure, and conventionally, we list logical forms like this: IAI-1 (the mood, a dash, then a number between 1 and 4 for the figure).

There are 4 figures and 64 moods. That gives us 256 possible logical forms. It turns out that only 15 of these are valid. We need a way to test them. It is to that task we now turn.

## The Venn Diagram Test for Validity

To test syllogistic forms for validity, we proceed in three steps:

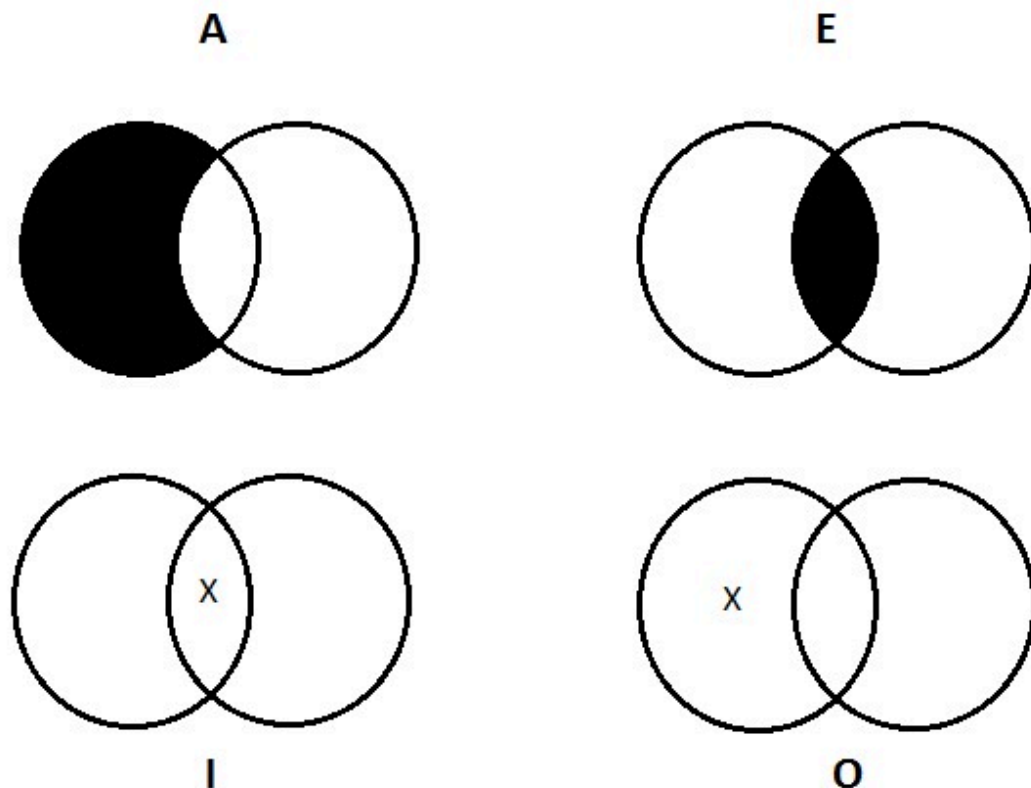
Draw three overlapping circles, like this:



That gives us one circle for each of the three terms in the syllogism: minor (S), major (P), and middle (M). It is useful to

write out your argument in standard form. Label the subject and predicate in the conclusion with S and P. Then match those labels to the other places those words appear in the premises. Write an M above the middle term shared by both premises.

Depict the assertions made by the premises of the syllogism on this diagram, using shading and Xs as appropriate, depicting the individual A, E, I, or O propositions in the usual way. For the sake of clarity, only the two relevant circles to any given premise are shown below – your Venn would have a third circle hanging out.



Each of the premises will be a proposition concerning only two of the three classes (S, P, and M). The major premise will concern M and P (in some order); the minor premise will concern M and S (in some order). How the circles will be labeled (with S, M, P) will depend on these particulars.

After the premises have been depicted on the three-circle diagram, we look at the finished product and ask, “Does this picture entail the truth of the conclusion?” If it does, the form is valid; if it does not, it is invalid.

In the course of running the test, we will keep two things in mind—one rule of thumb and one convention:

Rule of Thumb: If you can, go big first! In step 2, if a premise is an A or an E, start with that one. Depicting universal (A and E) premises before particular (I and O) ones (if there’s a choice) will help things go smoothly.

Convention: In cases of indeterminacy, draw Xs straddling boundary lines.

We need to explain what “indeterminacy” amounts to; we will in a moment. For now, to make all this more clear, we should run through some examples.

Let’s start at the beginning (alpha-numerically): AAA-1. We want to test this syllogistic form for validity. What does an argument of this form look like, schematically? Well, all three of its propositions are universal affirmatives, so they’re all of the form All \_\_ are \_\_. We have:

All \_\_ are \_\_

All \_\_ are \_\_

$\therefore$  All \_\_\_ are \_\_\_

That's what the mood (AAA) tells us. We have to figure out how to fill in the blanks with S, P, and M. The figure tells us how to do that. AAA-1: so, first figure. That looks like this:

(i) MP

SM

$\therefore$  SP

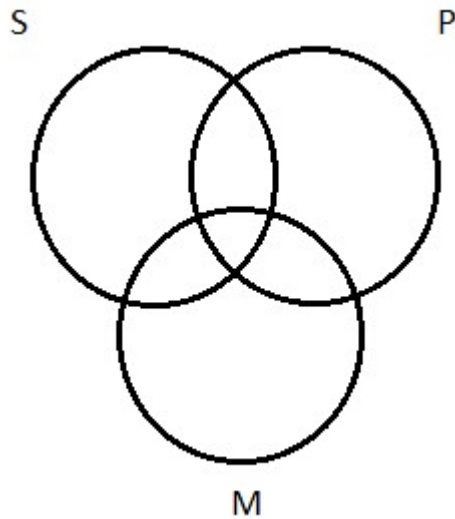
So AAA-1 can be schematically rendered thus:

All M are P.

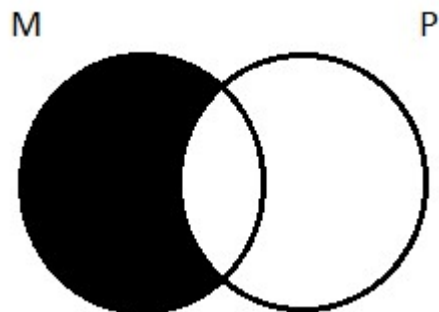
All S are M.

$\therefore$  All S are P.

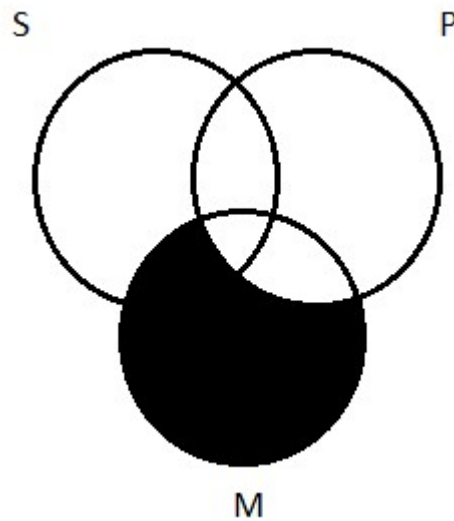
To test this form for validity, we start with step 1, and draw three circles:



In step 2, we depict the premises on this diagram. (We're supposed to keep in mind the rule of thumb that, given a choice, we should depict universal premises before particular ones, but since both of the premises are universals, this rule does not apply to this case.) We can start with the major premise: All M are P. On a regular two-circle Venn diagram, that would look like this:

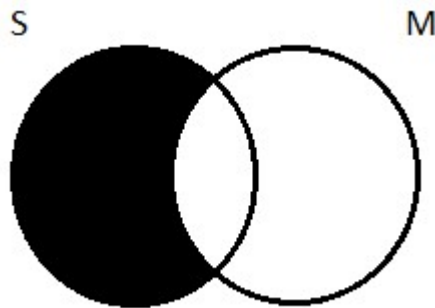


The trick is to transfer this two-circle diagram onto the three-circle one. In doing so, we keep in mind that all the parts of M that are outside of P must be shaded. That gives us this:

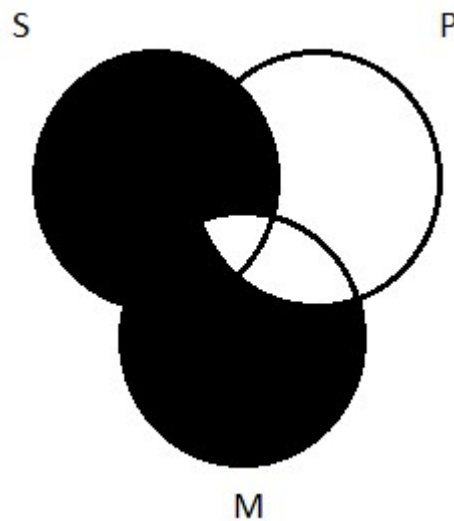


Note that in the course of shading out the necessary regions of M, we shaded out part of S. That's OK. Those members of the S class are Ms that aren't Ps; there's no such thing, so they have to go.

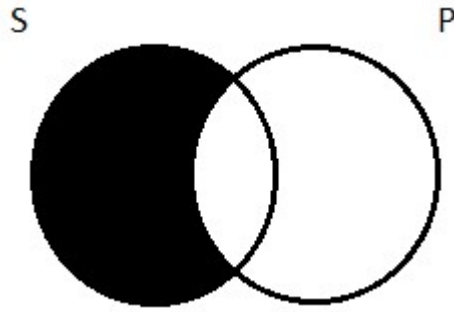
Next, we depict the minor premise: All S are M. With two circles, that would look like this:



Transferring that onto the three circle diagram means shading all the parts of S outside of M:



Step 2 is complete: we have depicted the assertions made by the premises. In step 3 we ask whether this diagram guarantees the truth of the conclusion. Well, our conclusion is All S are P. In a two-circle diagram, that looks like this:



Does our three-circle diagram guarantee the truth of All S are P? Focusing on the S and P circles, and comparing the two diagrams, there's a bit of a difference: part of the area of overlap between S and P is shaded out in our three-circle diagram, but it isn't in the two-circle depiction. But that doesn't affect our judgment about whether the diagram guarantees All S are P. Remember, this can be thought of as a claim that a certain kind of thing doesn't exist—an S that's outside the P circle. If there are any Ss (and there may not be), they will also be Ps. Our three-circle diagram does in fact guarantee this. There can't be an S that's not a P; those areas are shaded out. Any S you find will also be a P; it'll be in that little region in the center where all three circles overlap.

So, since the answer to our question is “yes”, the syllogistic form AAA-1 is valid. Trivial fact: all the valid syllogistic forms were given mnemonic nicknames in the Middle Ages to help students remember them. AAA-1 is called “Barbara”. No really. All the letters in the name had some meaning: the vowels indicate the mood (AAA); the other letters stand for features of the form that go beyond our brief investigation into Aristotelian Logic.

We should reflect for a moment on why this method works. We draw a picture that depicts the assertions made by the premises of the argument. Then we ask whether that picture guarantees the conclusion. This should sound familiar. We're testing for validity, and by definition, an argument is valid just in case its premises guarantee its conclusion; that is, IF the premises are true, then the conclusion must also be true. Our method mirrors the definition. When we depict the premises on the three-circle diagram, we're drawing a picture of what it looks like for the premises to be true. Then we ask, about this picture—which shows a world in which the premises are true—whether it forces us to accept the conclusion—whether it depicts a world in which the conclusion must be true. If it does, the argument is valid; if it doesn't, then it isn't. The method follows directly from the definition of validity.

To further illustrate the method, we should do some more examples. AII-3 is a useful one. The mood tells us it's going to look like this:

All \_\_ are \_\_  
Some \_\_ are \_\_  
∴ Some \_\_ are \_\_

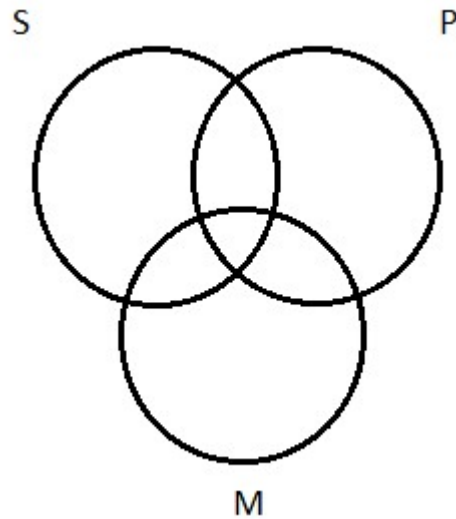
And we're in the third figure:

(iii) MP  
MS  
SP

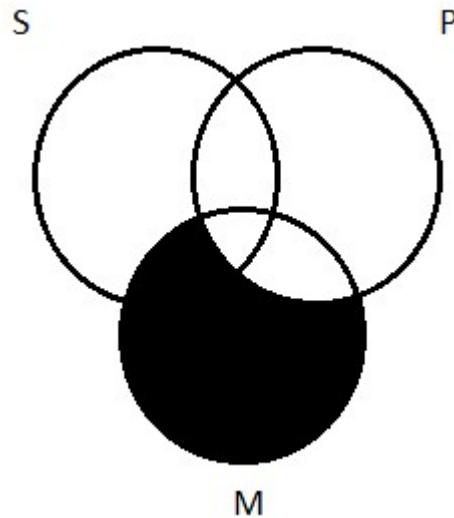
So we fill in the blanks to get the schematic form:

All M are P  
Some M are S  
∴ Some S are P

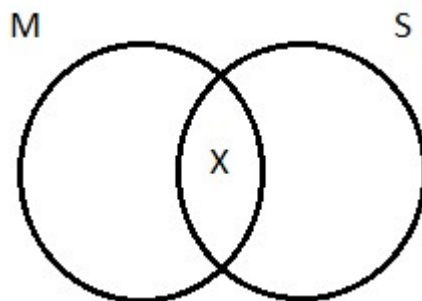
We start the test of this form with the blank three-circle diagram:



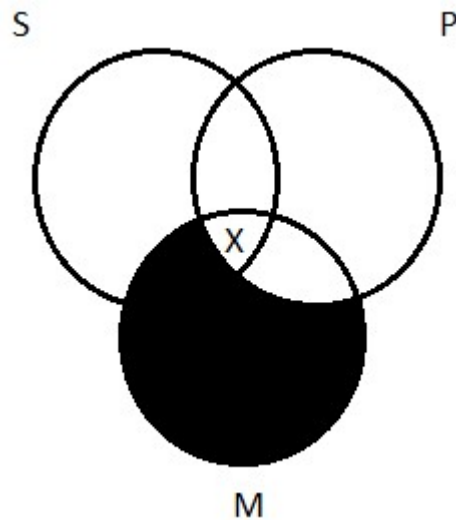
Step 2: depict the premises. And here, our rule of thumb applies: depict universals before particulars. The major premise is a universal (A) proposition; the minor premise is a particular (I). So we depict the major premise first. That's All M are P. We did this already. Recall that Barbara has the same major premise. So depicting that on the diagram gives us this:



Next, the minor premise: Some M are S. Recall, with particular propositions, we depict them using an X to indicate the thing said to exist. This proposition asserts that there is at least one thing that is both M and S:

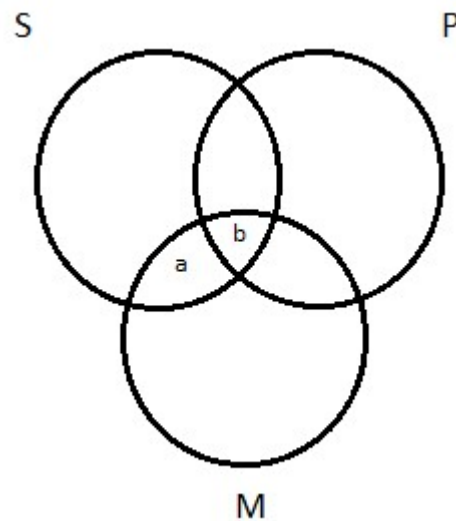


We need to transfer this to the three-circle diagram. We need an X that is in both the M and S circles. If we look at the area of overlap between the two, we see that part of it has been shaded out as the result of depicting the major premise, so there's only one place for the X to go:



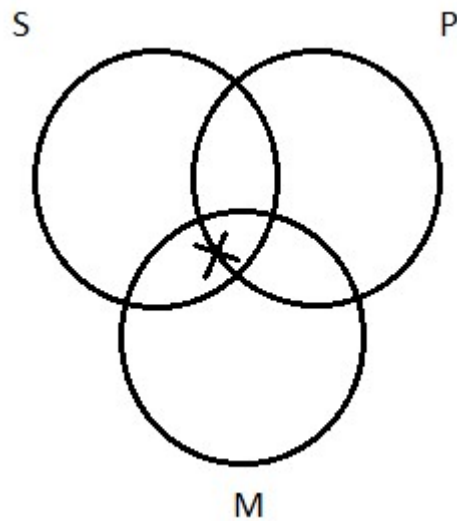
Step 2 is complete: the premises are depicted. So we proceed to step 3 and ask, “Does this picture guarantee the conclusion?” The conclusion is Some S are P; that’s an assertion that there is at least one thing that is both S and P. Is there? Yes! That X that we drew in the course of depicting the minor premise is in the sweet spot—the area of overlap between S and P. It guarantees the conclusion. The argument is valid. (If you’re curious, its mnemonic nickname is ‘Datisi’. Weird, I know; it was the Middle Ages.)

That’s another successful use of the Venn diagram test for validity, but I want to go back and revisit some of it. I want us to reflect on why we have the rule of thumb to depict universal premises before particular ones. Remember, we had the universal major premise All M are P and the particular minor premise Some M are S. The rule of thumb had us depict them in that order. Why? What would have happened had we done things the other way around? We would have started with a blank three-circle diagram and had to depict Some M are S on it. That means an X in the area of overlap between M and S. That area, though, is divided into two sub-regions (labeled ‘a’ and ‘b’):

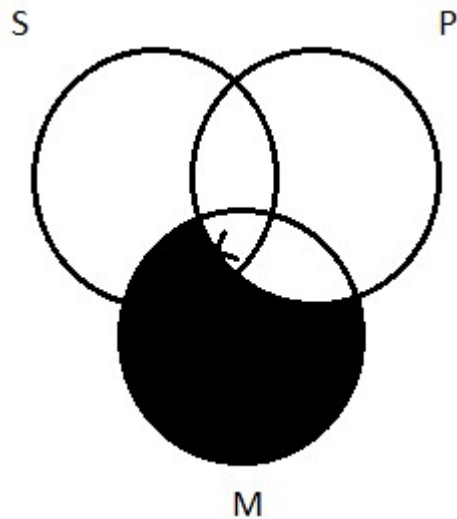


Where do I put my X—in region a or b? Notice, it makes a difference: if I put the X in region a, then it’s outside the P circle; if I put it in region b, then it’s inside the P circle. The question is: “Is this thing that the minor premise says exists a P or not a P?” I’m depicting a premise that only asserts Some M are S. That premise says nothing about P. It’s silent on our question; it gives us no guidance about how to choose between regions a and b. What to do? This is one of the cases of “indeterminacy” that we mentioned earlier when we introduced a convention to keep in mind when running the test for validity: In cases of indeterminacy, draw Xs straddling boundary lines. We don’t have any way of choosing between regions a and b, so when we draw our X, we split the difference:

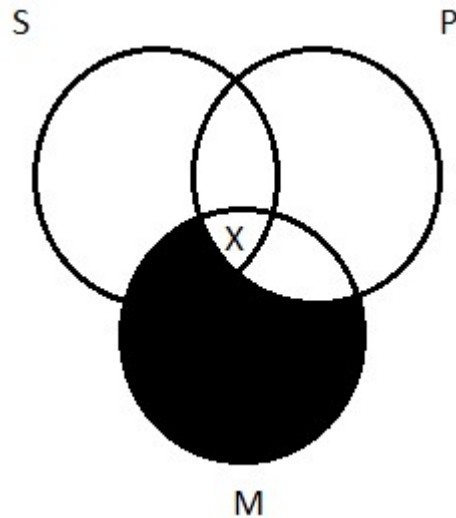




This drawing indicates that there's an X in there somewhere, either inside or outside the P circle, we don't know which. And now we see the reason for our rule of thumb—depict universals before particulars. Because if we proceed to depict the universal premise All M are P, we shade thus:



The shading erased half our X. That is, it resolved our question of whether or not the X should go in the P circle: it should. So now we have to go back and erase the half-an-X that's left and re-draw the X in that center region and end up with the finished diagram we arrived at earlier:

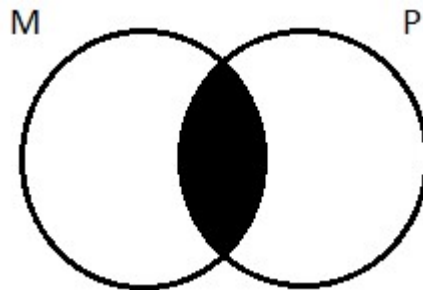


We would've saved ourselves the trouble had we just followed the rule of thumb to begin with and depicted the universal before the particular—shading before the X. That's the utility of the rule: sometimes it removes indeterminacy that would otherwise be present.

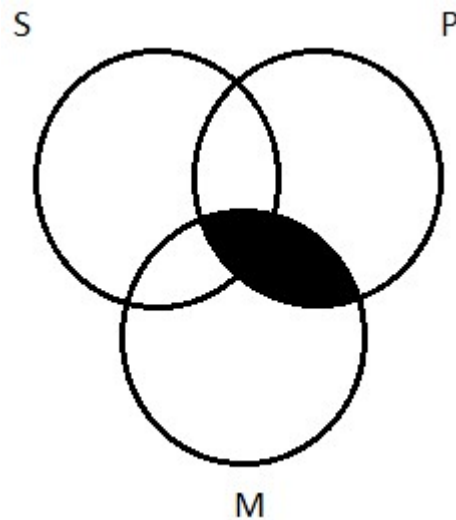
One more example to illustrate how this method works. Let's test EOI-1. Noting that in the first figure the middle term is first subject and then predicate, we can quickly fill in the schema:

No M are P  
 Some S are not M.  
 $\therefore$  Some S are P.

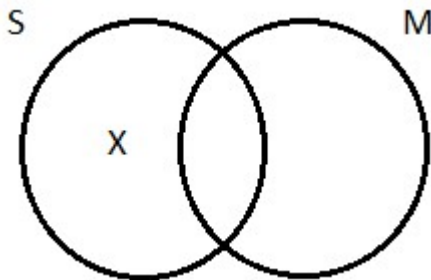
Following the rule of thumb, we depict the universal (E) premise first. No M are P asserts that there is nothing that is in both of those classes. The area of overlap between them is empty. With two circles, we have this:



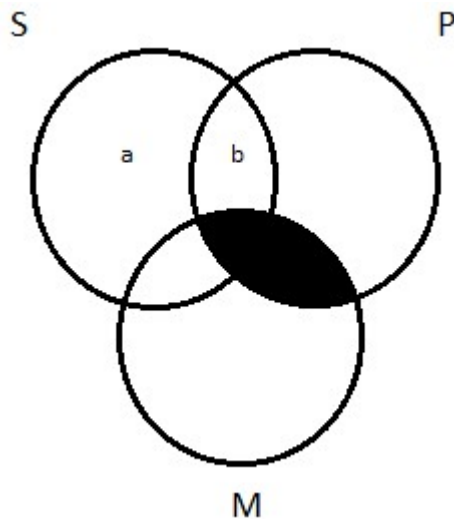
Transferring this onto the three-circle diagram, we shade out all the area of overlap between the M and P circles (clipping off part of S along the way):



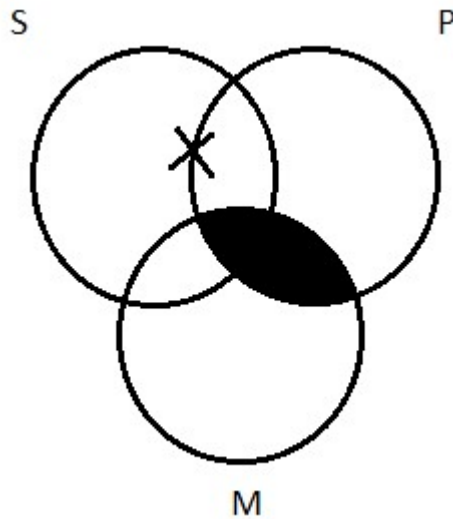
Next, the particular (O) premise: Some S are not M. This asserts the existence of something— namely, a thing that is an S but not an M. We need an X in the S circle that outside the M circle:



Moving to the three-circle diagram, though, things get messy. The area of S that's outside of M is divided into two sub-regions (labeled 'a' and 'b'):



We need an X somewhere in there, but do we put it in region a or region b? It makes a difference: if we put it in region b, it is a P; if we put it in region a, it is not. This is the same problem we faced before. We're depicting a premise—Some S are not M—that is silent on the question of whether or not the thing is a P. Indeterminacy. We can't decide between a and b, so we split the difference:



That X may be inside of P, or it may not; we don't know. This is a case in which we followed the rule of thumb, depicting the universal premise before the particular one, but it didn't have the benefit that it had when we tested AII-3: it didn't remove indeterminacy. That can happen. The rule of thumb is in place because it *sometimes* removes indeterminacy; it doesn't always work, though.

So now that we've depicted the premises, we ask whether they guarantee the conclusion. Is the world depicted in our diagram one in which the conclusion must be true? The conclusion is Some S are P: it asserts that there is at least one thing that is both S and P. Does our picture have such a thing? There's an X in the picture. Does it fit the bill? Is it both S and P? Well, uh... Maybe? That X may be inside of the area of overlap between S and P; then again, it may not be.

Oy. What do we say? It's tempting to say this: we don't know whether the argument is valid or not; it depends on where that X really is. But that's not the correct response. Remember, we're testing for validity—for whether or not the premises *guarantee* the conclusion. We can answer that question: they don't. For a guarantee, we would need an X in our picture that is *definitely* inside that middle region. We don't have such an X. These premises leave open the possibility that the conclusion is true; they don't rule it out. But that's not enough for validity. For an argument to be valid, the premises must *necessitate* the conclusion, force it on us. These do not. The form EOI1 is not valid.

## EXERCISES

1. Identify the logical form of the following arguments.

(a) Because some Wisconsinites are criminals and all criminals are scoundrels, it follows that some scoundrels are Wisconsinites.

(b) No surfers are priests, because all priests are men and some surfers are not men.

(c) Some authors are feminists, since some women are authors and some women are feminists.

(d) All mosquitoes are potential carriers of disease; therefore some mosquitoes are a menace to society, since all potential carriers of disease are a menace to society.

(e) Because some neo-Nazis are bloggers, some neo-Nazis are not geniuses, since no geniuses are bloggers.

2. Test the following syllogistic forms for validity.

(a) EAE-2

(b) EAE-3

(c) OAO-3

(d) EIO-4

(e) AOO-4

(f) IAI-1

(g) AII-1

3. Test the following arguments for validity.

(a) Some pirates are mercenaries; hence, some sailors are pirates, because all sailors are mercenaries.

(b) Some women are not nuns, but all nuns are sweethearts; it follows that some women are not sweethearts.

(c) Some forks are not utensils, for some forks are not cutlery, and some cutlery are not utensils.

4. Test the arguments in Exercise 1 for validity.

## Deductive Logic II: Propositional Logic

### I. Why Another Method of Deductive Logic?

Aristotle's syllogistic logic was great. It had a two-plus millennium run as the only game in town. As recently as the late 18<sup>th</sup> century (remember, Aristotle did his work in the 4<sup>th</sup> century BCE), the great German philosopher Immanuel Kant remarked that "since the time of Aristotle [logic] has not had to go a single step backwards... [and] it has also been unable to take a single step forward, and therefore seems to all appearance to be finished and complete."<sup>1</sup>

That may have been the appearance in Kant's time, but only because of an accident of history. In his own time, in ancient Greece, Aristotle's system had a rival—the logic of the Stoic school, culminating in the work of Chrysippus. Recall, for Aristotle, the fundamental logical unit was the class; and since terms pick out classes, his logic is often referred to as a "term logic". For the Stoics, the fundamental logical unit was the proposition; we could call this "propositional logic". These two approaches to logic were developed independently. Because of the vicissitudes of intellectual history (more later commentators promoted Aristotelian Logic, original writings from Chrysippus didn't survive, etc.), it turned out that Aristotle's approach was the one passed on to future generations, while the Stoic approach lay dormant. However, in the 19<sup>th</sup> century, thanks to work by logicians like George Boole (and many others), the propositional approach was revived and developed into a formal system.

Why is this alternative approach valuable? One of the concerns we had when we were introducing Aristotelian Logic was that, because of the restriction to categorical propositions, we would be limited in the number and variety of actual arguments we could evaluate. We brushed aside these concerns with a (somewhat vague) promise that, as a matter of fact, lots of sentences that were not standard form categoricals could be translated into that form. Furthermore, the restriction to categorical syllogisms was similarly unproblematic (we assured ourselves), because lots of arguments that are not standard form syllogisms could be rendered as (possibly a series of) such arguments.

These assurances are true in a large number of cases. But there are some very simple arguments that resist translation into strict Aristotelian form, and for which we would like to have a simple method for judging them valid. Here is one example:

Either Allende will win the election or Tomic will win the election.

Tomic will not win the election.

/∴ Allende will win the election.

None of the sentences in this argument is in standard form. And while the argument has two premises and a conclusion, it is not a categorical syllogism. Could we translate it into that form? Well, we can make some progress on the second premise and the conclusion, noting, as we did in

Chapter 3, that there's a simple trick for transforming sentences with singular terms (names like 'Allende' and 'Pinochet') into categoricals: let those names be class terms referring to the unit class containing the person they refer to, then render the sentences as universal. So the conclusion, 'Allende will win the election' can be rewritten in standard form as 'All Allendes are election-winners', where 'Allendes' refers to the unit class containing only Salvador Allende. Similarly, 'Tomic will not win the election' could be rewritten as a universal negative: 'No Tomics are election-winners'. The first

1. Kant, I. 1997. Critique of Pure Reason. Guyer, P. and Wood, A. (tr.). Cambridge: Cambridge University Press. p. 106.

premise, however, presents some difficulty: how do I render an either/or claim as a categorical? What are my two classes? Well, election-winners is still in the mix, apparently. But what to do with Allende and Tomic? Here's an idea: stick them together into the same class (they're not gonna like this), a class containing just the two of them. Let's call the class 'candidates'. Then this universal affirmative plausibly captures the meaning of the original premise: 'All election-winners are candidates'. So now we have this:

All election-winners are candidates.  
No Tomics are election-winners.  
∴ All Allendes are election-winners.

At least all the propositions are now categoricals. The problem is, this is not a categorical syllogism. Those are supposed to involve exactly three classes; this argument has four—Allendes, Tomics, election-winners, and candidates. True, candidates is just a composite class made by combining Allendes and Tomics, so you can make a case that there are really only three classes here. But, in a categorical syllogism, each of the class terms is supposed to occur exactly twice. 'Election-winners' occurs in all three, and I don't see how I can eliminate one of those occurrences.

Ugh. This is giving me a headache. It shouldn't be this hard to analyze this argument. You don't have to be a logician (or a logic student who's made it through three chapters of this book) to recognize that the Allende/Tomic argument is a valid one. Pick a random person off the street, show them that argument, and ask them if it's any good. They'll say it is. It's easy for regular people to make such a judgment; shouldn't it be easy for a logic to make that judgment, too? Aristotle's logic doesn't seem to be up to the task. We need an alternative approach.

This particular example is exactly the kind of argument that begs for a proposition-focused logic, as opposed to a class-focused logic like Aristotle's. If we take whole propositions as our fundamental logical unit, we can see that the form of this argument—the thing, remember, that determines its validity—is something like this:

Either A or T  
Not T  
∴ A

In this schema, 'A' stands for the proposition that Allende will win and 'T' for the proposition that Tomic will win. It's easy to see that this is a valid form.<sup>2</sup> This is the advantage of switching to a sentential, rather than a term, logic. It makes it easy to analyze this and many other argument forms.

In this chapter, we will discuss the basics of the proposition-centered approach to deductive logic—Propositional Logic. As was the case with Aristotelian Logic, Propositional Logic must accomplish three tasks:

1. Tame natural language.
2. Precisely define logical form.
3. Develop a way to test logical forms for validity.

The approach to the first task—taming natural language—will differ substantially from Aristotle's. Whereas Aristotelian Logic worked within a well-behaved portion of natural language—the sentences expressing categorical propositions—Propositional Logic constructs an artificial language and evaluates arguments expressed in its formalized terms. This move, of course, raises the concern we had about the applicability to everyday arguments even more acutely: what good is a logic if it doesn't evaluate English arguments at all? What we must show to alleviate this concern is that there is a systematic relationship between our artificial language and our natural one (English); we must show how

2. This form is often called the "Disjunctive Syllogism" and will be covered in Chapter 5. For now, notice that the word 'syllogism' is used there. By the Middle Ages, Stoic Logic hadn't disappeared entirely. Rather, bits of it were simply added on the Aristotelian system. So, it was traditional (and still is in many logic textbooks), when discussing Aristotelian Logic, to present this form, along with some others, as additional valid forms (supplementing Barbara, Datisi, and the rest). But this conflation of the two traditions obscures the fundamental difference between a class-centered approach to logic focused on categories and one focused on propositions. These should be kept distinct.

to translate between the two—and how translating *from* English into the artificial language results in the removal of imprecision and unruliness, the taming of natural language.

We will call our artificial language “SL,” short for ‘sentential logic’ as our goal is translating the propositions expressed by sentences into a formal language. There can be various formal languages designed to work with propositional logic. Each might have variations in the symbols it uses or the rules it apply to govern how those symbols function. SL is our choice because it is one of the relatively symbol options.

In constructing a language, we must specify its *syntax* and its *semantics*. The syntax of a language is the rules governing what counts as a well-formed construction within that language; that is, syntax is the language’s grammar. Syntax is what tells me that ‘What a handsome poodle you have there.’ is a well-formed English construction, while ‘Poodle a handsome there you what have.’ is not. The semantics of a language is an account of the meanings of its well-formed bits. If you know what a sentence means, then you know what it takes for it to express a truth or a falsehood. So semantics tells you under what conditions a given proposition is true or false.<sup>3</sup> Our discussion of the semantics of SL will reveal its relationship to English and tell us how to translate between the two languages.

## II. Syntax of SL

First, we cover syntax. This discussion will give us some clues as to the relationship between SL and English, but a full accounting of that relationship will have to wait, as we said, for the discussion of semantics.

We can distinguish, in English, between two types of (declarative) sentences: *simple* and *compound*. A simple sentence is one that does not contain any other sentence as a component part. A compound sentence is one that contains at least one other sentence as a component part. (We will not give a rigorous definition of what it is for one sentence to be a *component part* of another sentence. Rather, we will try to establish an intuitive grasp of the relation by giving examples, and stipulate that a rigorous definition *could* be provided, but is too much trouble to bother with.) ‘Beyoncé is logical’ is a simple sentence; none of its parts is itself a sentence.<sup>4</sup> Because it is a simple sentence that can’t be broken down, we say it expresses an *atomic proposition*. ‘Beyoncé is logical and James Brown is alive’ is a compound sentence: it contains two simple sentences as component parts—namely, ‘Beyoncé is logical’ and ‘James Brown is alive’. Compound sentences like this express *compound propositions*.

In SL, we will use capital letters—‘A’, ‘B’, ‘C’, ..., ‘Z’—to stand for the idea expressed by simple sentences. These are called *propositional constants*. Our practice will be simply to choose capital letters for simple sentences that are easy to remember. For example, we can choose ‘B’ to stand for ‘Beyoncé is logical’ and ‘J’ to stand for ‘James Brown is alive’. Easy enough. In Chapter 6 we will add a layer of detail, but for now we can use propositional constants to stand for *that which is expressed by a declarative sentence*, i.e., propositions. The hard part is symbolizing compound sentences in SL as compound propositions. How would we handle ‘Allende is elected or Tomic is elected’, for example? If we convert the entire sentence into a constant, E for example, we lose the information that helped us work through the Allende/Tomic

3. That’s actually a controversial claim about the role of semantics. Your humble author [Knachel], for example, is one of the weirdos who thinks it not true (of natural language, at least). But let’s leave those deviant linguists and philosophers (and their abstruse arguments) to one side and just say: semantics gives you truth-conditions. That’s certainly true of our artificial language SL.
4. You might think ‘Beyoncé is’ is a part of the sentence that qualifies as a sentence itself—a sentence claiming that she exists, maybe. But that won’t do. The word ‘is’ in the original sentence is the “‘is’ of predication”—a mere linking verb; ‘Beyoncé is’ only counts as a sentence if you change the meaning of ‘is’ to the “‘is’ of existence”. Anyway, stop causing trouble. This is why we didn’t give a rigorous definition of ‘component part’; we’d get bogged down in these sorts of arcane distinctions.



argument. Better to preserve the simple parts of the sentence and combine them in a functional way. Well, we've got capital letters to stand for the simple parts of the sentence, but that leaves out the word 'and'. We need more symbols.

## Logical Operators

We will distinguish five different kinds of compound propositions, and introduce a special SL symbol for each. Again, at this stage we are only discussing the syntax of SL—the rules for combining its symbols into well-formed constructions. We will have some hints about the semantics of these new symbols—hints about their meanings—but a full treatment of that topic will not come until the next section.

Our new symbols are called *logical operators*. Consider how the word “not” functions in a sentence; if we insert a “not” into “James Brown is alive” we get the reverse of the original proposition, i.e., James Brown is not alive. Thus, in the following section we will introduce a logical operator that can function as “not” and further help SL tame that part of language.

Other logical operators deal with compound sentences and the relationship between two simple propositions. Consider how A and T might be combined by using them to fill in the blanks around the following terms: “if \_ then \_,” “\_ and \_,” “\_ or \_” plus finally, “\_ if and only if \_.” In the following sections we will explore the logical operators that function in SL like how “and,” “or,” “not,” “if . . . then . . .” and “if and only if” do in English.

### Conjunctions

The first type of compound sentence is one that we've already seen. Conjunctions are, roughly, ‘and’-sentences—sentences like ‘Beyoncé is logical and James Brown is alive’. We've already decided to let ‘B’ stand for ‘Beyoncé is logical’ and to let ‘J’ stand for ‘James Brown is alive’. What we need is a symbol that stands for ‘and’. Many of us have seen & in a sentence. Well, instead of &, in SL, that symbol is a “dot”. It looks like this: •.

To form a conjunction in SL, we simply stick the dot between the two component letters, thus:

$$B \bullet J$$

That is the SL version of ‘Beyoncé is logical and James Brown is alive’.

A note on terminology. A conjunction has two components, one on either side of the dot. We will refer to these as the “*conjuncts*” of the conjunction. If we need to be specific, we might refer to the “left-hand conjunct” (‘B’ in this case) or the “right-hand conjunct” (‘J’ in this case).

### Disjunctions

Disjunctions are, roughly, ‘or’-sentences—sentences like ‘Beyoncé is logical or James Brown is alive’. Sometimes, the ‘or’ is accompanied by the word ‘either’, as in ‘Either Beyoncé is logical or James Brown is alive’. Again, we let ‘B’ stand for ‘Beyoncé is logical’ and let ‘J’ stand for ‘James Brown is alive’. What we need is a symbol that stands for ‘or’ (or ‘either/or’). In SL, that symbol is a “wedge”. It looks like this:  $\vee$ .

To form a disjunction in SL, we simply stick the wedge between the two component letters, thus:

$$B \vee J$$

That is the SL version of ‘Beyoncé is logical or James Brown is alive’.

A note on terminology. A disjunction has two components, one on either side of the wedge. We will refer to these as the

“disjuncts” of the disjunction. If we need to be specific, we might refer to the “left-hand disjunct” (‘B’ in this case) or the “right-hand disjunct” (‘J’ in this case).

In SL, disjunctions are always inclusive. This means that when we are given  $B \vee J$  we should understand this as saying B is true or J is true, or *both*. One reason for treating disjunctions as inclusive is that as we enrich SL, it is easier to start with inclusion and specify exclusive relations in another way, for instance if we need to say things like “ $B \vee J$  and not B and J.”

## Negations

Negations are, roughly, ‘not’-sentences—sentences like ‘James Brown is not alive’. You may find it surprising that this would be considered a compound sentence. It is not immediately clear how any component part of this sentence is itself a sentence. Indeed, if the definition of ‘component part’ (which we intentionally have not provided) demanded that parts of sentences contain only contiguous words (words next to each other), you couldn’t come up with a part of ‘James Brown is not alive’ that is itself a sentence. But that is not a condition on ‘component part’. In fact, this sentence does contain another sentence as a component part—namely, ‘James Brown is alive’. This can be made more clear if we paraphrase the original sentence. ‘James Brown is not alive’ means the same thing as ‘It is not the case that James Brown is alive’. Now we have all the words in ‘James Brown is alive’ next to each other; it is clearly a component part of the larger, compound sentence. We have ‘J’ to stand for the simple component; we need a symbol for ‘it is not the case that’. In SL, that symbol is a “tilde”. It looks like this:  $\sim$ .

To form a negation in SL, we simply prefix a tilde to the simpler component being negated:

$$\sim J$$

This is the SL version of ‘James Brown is not alive’.

## Conditionals

Conditionals are, roughly, ‘if/then’ sentences—sentences like ‘If Beyoncé is logical, then James Brown is alive’. (James Brown is actually dead. But suppose Beyoncé is a “James Brown-truther”, a thing that I just made up. She claims that James Brown faked his death, that the Godfather of Soul is still alive, getting funky in some secret location. In that case, the conditional sentence might make sense.) Again, we let ‘B’ stand for ‘Beyoncé is logical’ and let ‘J’ stand for ‘James Brown is alive’. What we need is a symbol that stands for the ‘if/then’ part. In SL, that symbol is a “horseshoe”. It looks like this:  $\supset$ .

To form a conditional in SL, we simply stick the horseshoe between the two component letters (where the word ‘then’ occurs), thus:

$$B \supset J$$

That is the SL version of ‘If Beyoncé is logical, then James Brown is alive’.

A note on terminology. Unlike our treatment of conjunctions and disjunctions, we will distinguish between the two components of the conditional. The component to the left of the horseshoe will be called the “antecedent” of the conditional; the component after the horseshoe is its “consequent”. As we will see when we get to the semantics for SL, there is a good reason for distinguishing the two components.

## Biconditionals

Biconditionals are, roughly, ‘if and only if’-sentences—sentences like ‘Beyoncé is logical if and only if James Brown is alive’. (This is perhaps not a familiar locution. We will talk more about what it means when we discuss semantics.) Again,

we let 'B' stand for 'Beyoncé is logical' and let 'J' stand for 'James Brown is alive'. What we need is a symbol that stands for the 'if and only if' part. In SL, that symbol is a "triple-bar". It looks like this:  $\equiv$ .

To form a biconditional in SL, we simply stick the triple-bar between the two component letters, thus:

$$B \equiv J$$

That is the SL version of 'Beyoncé is logical if and only if James Brown is alive'.

There are no special names for the components of the biconditional.

## *Punctuation – Parentheses*

Our language, SL, is quite austere: so far, we have only 31 different symbols—the 26 capital letters, and the five symbols for the five different types of compound sentence. We will now add two more: the left- and right-hand parentheses. And that'll be it.

We use parentheses in SL for one reason (and one reason only): to remove ambiguity. To see how this works, it will be helpful to draw an analogy between SL and the language of simple arithmetic. The latter has a limited number of symbols as well: numbers, signs for the arithmetical operations (addition, subtraction, multiplication, division), and parentheses. The parentheses are used in arithmetic for disambiguation. Consider this combination of symbols:

$$2 + 3 \times 5$$

As it stands, this formula is ambiguous. I don't know whether this is a sum or a product; that is, I don't know which operator—the addition sign or the multiplication sign—is the main operator. We can use parentheses to disambiguate, and we can do so in two different ways:

$$(2 + 3) \times 5$$

or

$$2 + (3 \times 5)$$

And of course, where we put the parentheses makes a big difference. The first formula is a product; the multiplication sign is the main operator. It comes out to 25. The second formula is a sum; the addition sign is the main operator. And it comes out to 17. Different placement of parentheses, different results.

This same sort of thing is going to arise in SL. We use the same term we use to refer to the addition and multiplication signs—'operator'—to refer to dot, wedge, tilde, horseshoe, and triple-bar. (As we will see when we look at the semantics for SL, this is entirely proper, since the SL operators stand for mathematical functions on truth-values.) There are ways of combining SL symbols into compound formulas with more than one operator; and just as is the case in arithmetic, without parentheses, these formulas would be ambiguous. Let's look at an example.

Consider this sentence: 'If Beyoncé is logical and James Brown is alive, then I'm the Queen of England'. This is a compound sentence, but it contains both the word 'and' and the 'if/then' construction. And it has *three* simple components: the two that we're used to by now about Beyoncé and James Brown, which we've been symbolizing with 'B' and 'J', respectively, and a new one—'I'm the Queen of England'—which we may as well symbolize with a 'Q'. Based on what we already know about how SL symbols work, we would render the sentence like this:

$$B \bullet J \supset Q$$

But just as was the case with the arithmetical example above, this formula is ambiguous. I don't know what kind of compound sentence this is—a conjunction or a conditional. That is, I don't know which of the two operators—the dot or the horseshoe—is the main operator. In order to disambiguate, we need to add some parentheses. There are two ways this can go, and we need to decide which of the two options correctly captures the meaning of the original sentence:

$$(B \bullet J) \supset Q$$

or

$$B \bullet (J \supset Q)$$

The first formula is a conditional; horseshoe is its main operator, and its antecedent is a compound sentence (the conjunction 'B • J'). The second formula is a conjunction; dot is its main operator, and its right-hand conjunct is a

compound sentence (the conditional ' $J \supset Q$ '). We need to decide which of these two formulations correctly captures the meaning of the English sentence 'If Beyoncé is logical and James Brown is alive, then I'm the Queen of England'.

The question is, what kind of compound sentence is the original? Is it a conditional or a conjunction? It is not a conjunction. Conjunctions are, roughly (again, we're not really doing semantics yet), 'and'-sentences. When you utter a conjunction, you're committing yourself to both of the conjuncts. If I say, "Beyoncé is logical and James Brown is alive," I'm telling you that both of those things are true. If we construe the present sentence as a conjunction, properly symbolized as ' $B \bullet (J \supset Q)$ ', then we take it that the person uttering the sentence is committed to both conjuncts; she's telling us that two things are true: (1) Beyoncé is logical *and* (2) if James Brown is alive then she's the Queen of England. So, if we take this to be a conjunction, we're interpreting the speaker as committed to the proposition that Beyoncé is logical. But clearly she's not. She uttered 'If Beyoncé is logical and James Brown is alive, then I'm the Queen of England' to express dubiousness about Beyoncé's logicity (and James Brown's status among the living). This sentence is not a conjunction; it is a conditional. It's saying that *if* those two things are true (about Beyoncé and James Brown), then I'm the Queen of England. The utterer doubts both conjuncts in the antecedent. The proper symbolization of this sentence is the first one above: ' $(B \bullet J) \supset Q$ '.

Again, in SL, parentheses have one purpose: to remove ambiguity. We only use them for that. This kind of ambiguity arises in formulas, like the one just discussed, involving multiple instances of the operators dot, wedge, horseshoe, and triple-bar.

Notice that I didn't mention the tilde there. Tilde is different from the other four. Dot, wedge, horseshoe, and triple-bar are what we might call "two-place operators". There are two simpler components in conjunctions, disjunctions, conditionals, and biconditionals. Negations, on the other hand, have only one simpler component; hence, we might call tilde a "one-place operator". It only operates on one thing: the sentence it negates.

This distinction is relevant to our discussion of parentheses and ambiguity. We will adopt a convention according to which the tilde negates the first well-formed SL construction immediately to its right. This convention will have the effect of removing potential ambiguity without the need for parentheses. Consider the following combination of SL symbols:

$$\sim A \vee B$$

It may appear that this formula is ambiguous, with the following two possible ways of disambiguating:

$$\sim (A \vee B)$$

or

$$(\sim A) \vee B$$

But this is not the case. Given our convention—tilde negates the first well-formed SL construction immediately to its right—the original formula—' $\sim A \vee B$ '—is *not* ambiguous; it is well-formed. Since 'A' is itself a well-formed SL construction (of the simplest kind), the tilde in ' $\sim A \vee B$ ' negates the 'A' only. This means that we don't have to indicate this fact with parentheses, as in the second of the two potential disambiguations above. That kind of formula, with parentheses around a tilde and the item it negates, is not a well-formed construction in SL. Given our convention about tildes, the parentheses around ' $\sim A$ ' are redundant.

The first potential disambiguation—' $\sim (A \vee B)$ '—is well-formed, and it means something different from ' $\sim A \vee B$ '. In the former, the tilde negates the entire disjunction, ' $A \vee B$ '; in the latter, it only negates 'A'. That makes a difference. Again, an analogy to arithmetic is helpful here. Compare the following two formulas:

$$-(2 + 5)$$

vs.

$$-2 + 5$$

In the first, the minus-sign covers the entire sum, and so the result is -7; in the second, it only covers the 2, so the result is 3. This is exactly analogous to the difference between ' $\sim (A \vee B)$ ' and ' $\sim A \vee B$ '. The tilde has wider scope in the first formula, and that makes a difference. The difference can only be explained in terms of meaning—which means it is time to turn our attention to the semantics of SL.

### III. Semantics of SL

Our task is to give precise meanings to all of the well-formed formulas of SL. We will refer to these, quite sensibly, as “sentences of SL” or “statements in SL.” Some of this task is already complete. We know something about the meanings to the 26 capital letters: they stand for simple propositions of our choosing. While the semantics for a natural language like English is complicated, the semantics for SL sentences is simple: all we care about is truth-value. A sentence in SL can have one of two semantic values: true or false. That’s it. And so when we ask, “What is the exact meaning of a statement in SL,” we find the answer by answering the question, under what circumstances do the logical operators express true statements, i.e, what are their “truth-conditions”? For example, when is  $A \supset B$  true?, and when false? If A is false, but B true, what is the *truth value* of the  $\supset$  and the entire compound statement? Well, the truth value is determined by the “truth conditions” for each operator. This section will explain this idea further and provide the truth conditions for our new logical operators.

This is one of the ways in which the move to SL is a taming of natural language. In SL, every sentence has a determinate truth-value; and there are only two choices: true or false. English and other natural languages are more complicated than this. Of course, there’s the issue of non-declarative sentences, which don’t express propositions and don’t have truth-values at all. But even if we restrict ourselves to declarative English sentences, things don’t look quite as simple as they are in SL. Consider the sentence ‘Napoleon was short’. You may not be aware that the popular conception of the French Emperor as diminutive in stature has its roots in British propaganda at the time. As a matter of fact, he was about 5’ 7”.

The problem here is that relative terms like ‘short’ have borderline cases; they’re vague. It’s not clear how to assign a truth-value to sentences like ‘Napoleon is short’. So, in English, we might say that they lack a truth-value (absent some explicit specification of the relevant standards). Logics that are more sophisticated than our SL have developed ways to deal with these sorts of cases. Instead of just two truth-values, some logics add more. There are three-values logics, where you have true, false, and neither. So we could say ‘Napoleon is short’ is neither. There are logics with infinitely many truth-values between true and false (where false is zero and true is 1, and every real number in between is a degree of truth); in such a system, we could assign, I don’t know, .62 to the proposition that Napoleon is short. The point is, English and other natural languages are messy when it comes to truth-value. We’re taming them in SL by assuming that every SL sentence has a determinate truth-value, and that there are only two truth-values: true and false—which we will indicate, by the way, with the letters ‘T’ and ‘F’.

Our task from here is to provide semantics for the five operators:  $\sim$  tilde,  $\bullet$  dot,  $\vee$  wedge,  $\equiv$  triple-bar, and  $\supset$  horseshoe (we start with tilde because it’s the simplest, and we save horseshoe for last because it’s quite a bit more involved). We will specify the meanings of these symbols in terms of their effects on truth-value: what is the truth-value of a compound sentence featuring them as the main operator, given the truth-values of the components? The semantic values of the operators will be truth functions: systematic accounts of the truth-value outputs (of the compound sentence) resulting from the possible truth-value inputs (of the simpler components).

As we talk about SL, we will use lower case letters, such as “p” and “q” to stand in as *variables* which in turn might be replaced by any determined set of propositional constants. For example, while  $A \supset B$  says something about propositions A and B, “ $p \supset q$ ” merely expresses the general logical form which could later be filled in.

#### *Negations (TILDE) ~*

Because tilde is a one-place operator, this is the simplest operator to deal with. Again, the general form of a negation is  $\sim p$ , where ‘p’ is a variable standing for any generic SL sentence, simple or compound. As a lower-case letter, ‘p’ is not part of our language (SL); rather, it’s a tool we use to talk about our language—to refer to generic well-formed constructions within it.

We need to give an account of the meaning of the tilde in terms of its effect on truth-value. Tilde, as we said, is the SL

equivalent of ‘not’ or ‘it is not the case that’. Let’s think about what happens in English when we use those terms. If we take a true sentence, say ‘Edison invented the light bulb’, and form a compound with it and ‘not’, we get ‘Edison did not invent the light bulb’—a falsehood. If we take a false sentence, like ‘James Brown is alive’, and negate it, we get ‘James Brown is not alive’—a truth.

Evidently, the effect of negation on truth-value is to turn a truth into a falsehood, and a falsehood into a truth. We can represent this graphically, using what we’ll call a “truth-table.” The following table gives a complete specification of the semantics of tilde. In others, the *meaning* of  $\sim$  is represented in the truth-table.

p	$\sim p$
T	F
F	T

In the left-hand column, we have ‘p’, which, as a variable, stands for a generic, unspecified SL sentence. Since it’s unspecified, we don’t know its truth-value; but since it’s a sentence in SL, we do know that there are only two possibilities for its truth-value: true or false (T or F). So in the first column, we list those two possibilities. In the second column, we have ‘ $\sim p$ ’, the negation of whatever ‘p’ is. We can compute the truth-value of the negation based on the truth-value of the sentence being negated: if the original sentence is true, then its negation is false; if the original sentence is false, then the negation is true. This is what we represent when we write ‘F’ and ‘T’ underneath the tilde (the operator that effects the change in truth-value) in the second column, in the same rows as their opposites.

Tilde is a *logical operator*. Its meaning is specified by a function: if you input a T, the output is an F; if you input an F, the output is a T. The other four operators will also be defined in terms of the truth-function they represent. This is exactly analogous, again, to arithmetic. Addition, with its operator ‘+’, is a function on numbers. Input 1 and 3, and the output is 4. In SL, we only have two values—T and F—but it’s the same kind of thing. We could just as well use numbers to represent the truth-values: 0 for false and 1 for true, for example. In that case, tilde would be a function that outputs 0 when 1 is the input, and outputs 1 when 0 is the input.

### Conjunctions (DOT) •

Our rough-and-ready characterization of conjunctions was that they are ‘and’-sentences— sentences like ‘Beyoncé is logical and James Brown is alive’. Since these sorts of compound sentences involve two simpler components, we say that dot is a two-place operator. So when we specify the general form of a conjunction using generic variables, we need two of them. The general form of a conjunction in SL is  $p \bullet q$ . The questions we need to answer are these: Under what circumstances is the entire conjunction true, and under what circumstances false? And how does this depend on the truth-values of the component parts?

We remarked earlier that when someone utters a conjunction, they’re committing themselves to both of the conjuncts. If I tell you that Beyoncé is wise and James Brown is alive, I’m committing myself to the truth of both of those alleged facts; I am, as it were, promising you that both of those things are true. So, if even one of them turns out false, I’ve broken my promise; the only way the promise is kept is if both of them turn out to be true.

This is how conjunctions work, then: they’re true only when both conjuncts are true; false otherwise. We can represent this graphically, with a truth-table defining the dot:

p	q	$p \cdot q$
T	T	T
T	F	F
F	T	F
F	F	F

Since the dot is a two-place operator, we need columns for each of the two variables in its general form— $p$  and  $q$ . Each of these is a generic SL sentence that can be either true or false. That gives us four possibilities for their truth-values as a pair: both true,  $p$  true and  $q$  false,  $p$  false and  $q$  true, both false. These four possibilities give us the four rows of the table. For each of these possible inputs to the truth-function, we get an output, listed under the dot. T is the output when both inputs are Ts; F is the output in every other circumstance.

### *Disjunctions (WEDGE) $\vee$*

Our rough characterization of disjunctions was that they are ‘or’-sentences—sentences like ‘Beyoncé is logical or James Brown is alive’. In SL, the general form of a disjunction is  $p \vee q$ . We need to figure out the circumstances in which such a compound is true; we need the truth-function represented by the wedge.

At this point we face a complication. Wedge is supposed to capture the essence of ‘or’ in English, but the word ‘or’ has two distinct senses. This is one of those cases where natural language needs to be tamed: our wedge can only have one meaning, so we need to choose between the two alternative senses of the English word ‘or’.

‘Or’ can be used exclusively or inclusively. The exclusive sense of ‘or’ is expressed in a sentence like this: ‘King Kong will win the election or Godzilla will win the election’. The two disjuncts present exclusive possibilities: one or the other will happen, but not both. The inclusive sense of ‘or’, however, allows the possibility of both. If I told you I was having trouble deciding what to order at a restaurant, and said, “I’ll order lobster or steak,” and then I ended up deciding to get the surf ‘n’ turf (lobster and steak combined in the same entrée), you wouldn’t say I had lied to you when I said I’d order lobster or steak. The inclusive sense of ‘or’ allows for one or the other—or both.

We will use the inclusive sense of ‘or’ for our wedge. There are arguments for choosing the inclusive sense over the exclusive one, but we will not dwell on those here. We need to choose a meaning for wedge, and we’re choosing the inclusive sense of ‘or’. As we will see later, the exclusive sense will not be lost to us because of this choice: we will be able to symbolize exclusive ‘or’ within SL, using a combination of operators.

So, wedge is inclusive ‘or’. It’s true whenever one or the other—or both—conjuncts is true; false otherwise. This is its truth-table definition:

p	q	$p \vee q$
T	T	T
T	F	T
F	T	T
F	F	F

### Biconditionals (TRIPLE-BAR)

As we said, biconditionals are, roughly, ‘if and only if’-sentences—sentences like ‘Beyoncé is logical if and only if James Brown is alive’. ‘If and only if’ is not a phrase most people use in everyday life, but the meaning is straightforward: it’s used to claim that both components have the same truth-value, that one entails the other and *vice versa*, that they can’t have different truthvalues. In SL, the general form of a biconditional is  $p \equiv q$ . This is the truth-function:

p	q	$p \equiv q$
T	T	T
T	F	F
F	T	F
F	F	T

The triple-bar is kind of like a logical equals-sign (it even resembles ‘=’): the function delivers an output of T when both components are the same, F when they’re not. This is why we need not give specific names for each component of the biconditional – the components are interchangeable.

While the truth-functional meaning of triple-bar is now clear, it still may be the case that the intuitive meaning of the English phrase ‘if and only if’ remains elusive. This is natural. Fear not: we will have much more to say about that locution when we discuss translating between English and SL; a full understanding of *biconditionals* can only be achieved based on a full understanding of *conditionals*, to which, as the names suggest, they are closely related. We now turn to a specification of the truth-functional meaning of the latter.

### Conditionals (HORSESHOE) $\supset$

Our rough characterization of conditionals was that they are ‘if/then’ sentences—sentences like ‘If Beyoncé is logical, then James Brown is alive’. We use such sentences all the time in everyday speech, but is surprisingly difficult to pin down the precise meaning of the conditional, especially within the constraints imposed by SL. There are in fact many competing accounts of the conditional—many different conditionals to choose from—in a literature dating back all the way to the Stoics of ancient Greece. Whole books can be—and have been—written on the topic of conditionals. In the



course of our discussion of the semantics for horseshoe, we will get a sense of why this is such a vexed topic; it's complicated.

The general form of a conditional in SL is  $p \supset q$ . We need to decide for which values of  $p$  and  $q$  the conditional turns out true and false. To help us along (by making things more vivid), we'll consider an actual conditional claim, with a little story to go along with it. Suppose Barb is suffering from joint pain; maybe it's gout, maybe it's arthritis—she doesn't know and hasn't been to the doctor to find out. She's complaining about her pain to her neighbor, Sally. Sally is a big believer in “alternative medicine” and “holistic healing”. After hearing a brief description of the symptoms, Sally is ready with a prescription, which she delivers to Barb in the form of a conditional claim: “If you drink this herbal tea every day for a week, then your pain will go away.” She hands over a packet of tea leaves and instructs Barb in their proper preparation. We want to evaluate Sally's conditional claim—that if Barb drinks the herbal tea daily for a week, then her pain will go away—for truth/falsity. To do so, we will consider four various scenarios, the details of which will bear on that evaluation.

Scenario #1: Barb does in fact drink the tea every day for a week as prescribed, and, after doing so, lo and behold, her pain is gone. Sally was right! In this scenario, we would say that the conditional we're evaluating is true.

Scenario #2: Barb does as Sally said and drinks the tea every day for a week, but, after the week is finished, the pain remains, the same as ever. In this scenario, we would say that Sally was wrong: her conditional advice was false.

Perhaps you can see what I'm doing here. Each of the scenarios represents one of the rows in the truth-table definition for the horseshoe. We will swap our place-holder variables ( $p$  and  $q$ ) for propositional constants that stand for the claims being made in our example. Sally's conditional claim has an antecedent—Barb drinks the tea every day for a week—and a consequent—Barb's pain is relieved. These are  $D$  and  $R$ , respectively, in the conditional  $D \supset R$ . In scenario #1, both  $D$  and  $R$  were true: Barb did drink the tea, and the pain did go away; in scenario #2,  $D$  was true (Barb drank the tea) but  $R$  was false (the pain didn't go away). These two scenarios are the first two rows of the four-row truth tables we've already seen for dot, wedge, and triple-bar. For horseshoe, the truth-function gives us  $T$  in the first row and  $F$  in the second:

D	R	$D \supset R$
T	T	T
T	F	F

All that's left is to figure out what happens in the third and fourth rows of the table, where the antecedent ( $D$ , Barb drinks the tea) is false both times and the consequent is first true (in row 3) and then false (in row 4). There are two more scenarios to consider.

In scenario #3, Barb decides Sally is a bit of a nut, or she drinks the tea once and it tastes awful so she decides to stop drinking it—whatever the circumstances, Barb doesn't drink the tea for a week; the antecedent is false. But in this scenario, it turns out that after the week is up, Barb's pain has gone away; the consequent is true. What do we say about Sally's advice—if you drink the tea, the pain will go away—in this set of circumstances?

In scenario #4, again Barb does not drink the tea (false antecedent), and after the week is up, the pain remains (false consequent). What do we say about the Sally's conditional advice in this scenario?

It's tempting to say that in the 3<sup>rd</sup> and 4<sup>th</sup> scenarios, since Barb didn't even try Sally's remedy, we're not in a position to evaluate Sally's advice for truth or falsity. The hypothesis wasn't even tested. So, we're inclined to say ‘If you drink the tea, then the pain will go away’ is neither true nor false. But while this might be a viable option in English, it won't work in SL. We've made the simplifying assumptions that every SL sentence must have a truth-value, and that that the only two possibilities are true and false. We can't say it has no truth-value; we can't add a third value and call it “neither”. We have to put a  $T$  or an  $F$  under the horseshoe in the third and fourth rows of the truth table for that operator. Given this restriction, and given that we've already decided how the first two rows should work out, there are four possible ways of specifying the truth-function for horseshoe:

D	R	(1)	(2)	(3)	(4)
T	T	T	T	T	T
T	F	F	F	F	F
F	T	F	T	F	T
F	F	F	F	T	T

These are our only options (remember, the top two rows are settled; scenarios 1 and 2 above had clear results). Which one captures the meaning of the conditional best?

Option 1 is tempting: as we noted, in rows 3 and 4, Sally's hypothesis isn't even tested. If we're forced to choose between true and false, we might as well go with false. The problem with this option is that this truth-function—true when both components are true; false otherwise—is already taken. That's the meaning of dot. If we choose option 1, we make horseshoe and dot mean the same thing. That won't do: they're different operators; they should have different meanings. 'And' and 'if/then' don't mean the same thing in English, clearly.

Option 2 also has its charms. OK, we might say, in neither situation is Sally's hypothesis tested, but at least row 3 has something going for it, Sally-wise: the pain does go away. So let's say her conditional is true in that case, but false in row 4 when there still is pain. Again, this won't do. Compare the column under option 2 to the column under R. They're the same: T, F, T, F. That means the entire conditional,  $D \supset R$ , has the same meaning as its consequent, plain old R. Not good. The antecedent, D, makes no difference to the truth-value of the conditional in this case. But it should; we shouldn't be able to compute the truth-value of a two-place function without even looking at one of the inputs.

Option 3 is next. Some people find it reasonable to say that the conditional is false in row 3: there's something about the disappearance of the pain, despite not drinking the tea, that's incompatible with Sally's prediction. And if we can't put an F in the last row too (this is just option 1 again), then make it a T. But this fails for the same reason option 1 did: the truth-function is already taken, this time by the triple-bar. 'If and only if' is a much stronger claim than the mere 'if/then'; biconditionals must have a different meaning from mere conditionals.

That leaves option 4. This is the one we'll adopt, not least because it's the only possibility left. The conditional is true when both antecedent and consequent are true—scenario 1; it's false when the antecedent is true but the consequent false—scenario 2; and it's true whenever the antecedent is false—scenarios 3 and 4. Stepping out of our example and swapping D and R out for variables p and q, we can then express the general form of conditionals. Thus, this is the definition of horseshoe:

p	q	$p \supset q$
T	T	T
T	F	F
F	T	T
F	F	T

It's not ideal. The first two rows are quite plausible, but there's something profoundly weird about saying that the sentence 'If you drink the tea, then the pain will go away' is true whenever the tea is not drunk. Yet that is our only option. We can perhaps make it a bit more palatable by saying— as we did about universal categorical propositions with

empty subject classes—that while it's true in such cases, it's only true vacuously or trivially—true in a way that doesn't tell you about how things are in the world.

What can also help a little is to point out that while rows 3 and 4 don't make much sense for the Barb/Sally case, they do work for other conditionals involving other atomic propositions in the place of  $p$  and  $q$ . The horror author Stephen King lives in Maine (half his books are set there, it seems). Consider this conditional: 'If Stephen King is the Governor of Maine, then he lives in Maine'. While a prominent citizen, King is not Maine's governor, so the antecedent is false. He is, though, as we've noted, a resident of Maine, so the consequent is true. We're in row 3 of the truth-table for conditionals here. And intuitively, the conditional is true: he's not the governor, but *if he were*, he would live in Maine (governors reside in their states' capitals). And consider this conditional: 'If Stephen King is president of the United States, then he lives in Washington, DC'. Now both the antecedent (King is president) and the consequent (he lives in DC) are false: we're in row 4 of the table. But yet again, the conditional claim is intuitively true: *if he were* president, he would live in DC.

Notice the trick I pulled there: I switched from the so-called indicative mood (if he is) to the subjunctive (if he *were*). The truth of the conditional is clearer in the latter mood than the former. But this trick won't always work to make the conditional come out true in the third and fourth rows. Consider: 'If Stephen King were president of the United States, then he would live in Maine' and 'If Stephen King were Governor of Maine, then he would live in Washington, DC'. These are third and fourth row examples, respectively, but neither is intuitively true.

By now perhaps you are getting a sense of why conditionals are such a vexed topic in the history of logic. A variety of approaches, with attendant alternative logical formalisms, have been developed over the centuries (and especially in the last century) to deal with the various problems that arise in connection with conditional claims. The one we use in SL is the very simplest approach, the one with which to begin. As this is an introductory text, this is appropriate. You can investigate alternative accounts of the conditional if you extend your study of logic further.

### *Computing Truth-Values of Compound SL Sentences*

With the truth-functional definitions of the five SL operators in hand, we can develop a preliminary skill that will be necessary to deploy when the time comes to test SL arguments for validity. We need to be able to compute the truth-values of compound SL sentences, given the truth-values of their simplest parts (the simple sentences—capital letters). To do so, we must first determine what type of compound sentence we're dealing with—negation, conjunction, disjunction, conditional, or biconditional. This involves deciding which of the operators in the SL sentence is the *main* operator. We then compute the truth-value of the compound according to the definition for the appropriate operator, using the truth-values of the simpler components. If these components are themselves compound, we determine their main operators and compute accordingly, in terms of their simpler components—repeating as necessary until we get down to the simplest components of all, the capital letters. A few examples will make the process clear.

Let's suppose that  $A$  and  $B$  are true SL sentences. Consider this compound:

$$\sim A \vee B$$

What is its truth-value? To answer that question, we first have to figure out what kind of compound sentence we're dealing with. It has two operators—the tilde and the wedge. Which of these is the main operator; that is, do we have a negation or a disjunction? We answered this question earlier, when we were discussing the syntax of SL. Our convention with tildes is that they negate the first well-formed construction immediately to their right. In this case, ' $A$ ' is the first well-formed construction immediately to the right of the tilde, so the tilde negates it. That means wedge is the main operator; this is a disjunction, where the left-hand disjunct is  $\sim A$  and the right-hand disjunct is  $B$ . To compute the truth-value of the disjunction, we need to know the truth-values of its disjuncts. We know (by our starting stipulation) that  $B$  is true; we need to know the truth-value of  $\sim A$ . That's easy, since  $A$  is true,  $\sim A$  must be false. It's helpful to keep track of one's step-by-step computations like so:

T	T
$\sim A \vee B$	
F	

I've marked the truth-values of the simplest components, A and B, on top of those letters. Then, under the tilde, the operator that makes it happen, I write 'F' to indicate that the left-hand disjunct,  $\sim A$ , is false. Now I can compute the truth-value of the disjunction: the left-hand disjunct is false, but the right hand disjunct is true; this is row 3 of the wedge truth-table, and the disjunction turns out true in that case. I indicate this with a 'T' under the wedge, which I highlight (with boldface and underlining) to emphasize the fact that this is the truth-value of the whole compound sentence:

T	T
$\sim A \vee B$	
F	
<b><u>T</u></b>	

When we were discussing syntax, we claimed that adding parentheses to a compound like the last one would alter its meaning. We're now in a position to prove that claim. Consider this SL sentence (where A and B are again assumed to be true):

T	T
$\sim (A \vee B)$	

Now the main operator is the tilde: it negates the entire disjunction inside the parentheses. To discover the effect of that negation on truth-value, we need to compute the truth-value of the disjunction that it negates. Both A and B are true; this is the top row of the wedge truth-table— disjunctions turn out true in such cases:

T	T
$\sim (A \vee B)$	
T	

So the tilde is negating a truth, giving us a falsehood:

T	T
$\sim A \vee B$	
T	
<b><u>F</u></b>	

The truth-value of the whole is false; the similar-looking disjunction without the parentheses was true. These two SL sentences must have different meanings; they have different truth-values. It will perhaps be useful to look at one more example, this time of a more complex SL sentence. Suppose again that A and B are true SL simple sentences, and that X and Y are false SL simple sentences. Let's compute the truth-value of the following compound sentence:

$$\sim (A \bullet X) \supset (B \vee \sim Y)$$

As a first step, it's useful to mark the truth-values of the simple sentences:

$\sim (A \bullet X) \supset (B \vee \sim Y)$			
T	F	T	F

Now, we need to figure out what kind of compound sentence this is; what is the main operator? This sentence is a conditional; the main operator is the horseshoe. The tilde at the far left negates the first well-formed construction immediately to its right. In this case, that is  $(A \bullet X)$ .  $\sim (A \bullet X)$  is the antecedent of this conditional;  $(B \vee \sim Y)$  is the consequent. We need to compute the truth values of each of these before we can compute the truth-value of the whole compound.

Let's take the antecedent,  $\sim (A \bullet X)$  first. The tilde negates the conjunction, so before we can know what the tilde does, we need to know the truth-value of the conjunction inside the parentheses. Conjunctions are true just in case both conjuncts are true; in this case, A is true but X is false, so the conjunction is false, and its negation must be true:

T	F	T	F
$\sim (A \bullet X) \supset (B \vee \sim Y)$			
F			
T			

So the antecedent of our conditional is true. Let's look at the consequent,  $(B \vee \sim Y)$ . Y is false, so  $\sim Y$  must be true. That means both disjuncts, B and  $\sim Y$  are true, making our disjunction true:

T	F	T	F
$\sim (A \bullet X) \supset (B \vee \sim Y)$			
F		T	
T		T	

Both the antecedent and consequent of the conditional are true, so the whole conditional is true:

T	F	T	F
$\sim (A \bullet X) \supset (B \vee \sim Y)$			
F			
T			
T			

One final note: sometimes you only need partial information to make a judgment about the truthvalue of a compound sentence. Look again at the truth table definitions of the two-place operators:

p	q	$p \bullet q$	$p \vee q$	$p \supset q$	$p \equiv q$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	F	T	F	T
F	F	F	F	T	T

For three of these operators—the dot, wedge, and horseshoe—one of the rows is not like the others. For the dot: it only comes out true when both p and q are true, in the top row. For the wedge: it only comes out false when both p and q are false, in the bottom row. For the horseshoe: it only comes out false when p is true and q is false, in the second row.

Noticing this allows us, in some cases, to compute truth-values of compounds without knowing the truth-values of both

components. Suppose again that A is true and X is false; and let Q be a simple SL sentence the truth-value of which is a mystery to you (it has one, like all of them must; I'm just not telling you what it is). Consider this compound:

$$A \vee Q$$

We know one of the disjuncts is true; we don't know the truth-value of the other one. But we don't need to! A disjunction is only false when both of its disjuncts are false; it's true when even one of its disjuncts is true. A being true is enough to tell us the disjunction is true; Q doesn't matter.

Consider the conjunction:

$$X \bullet Q$$

We only know the truth-value of one of the conjuncts: X is false. That's all we need to know to compute the truth-value of the conjunction. Conjunctions are only true when both of their conjuncts are true; they're false when even one of them is false. X being false is enough to tell us that this conjunction is false.

Finally, consider these conditionals:

$$Q \supset A \text{ and } X \supset Q$$

They are both true. Conditionals are only false when the antecedent is true and the consequent is false; so they're true whenever the consequent is true (as is the case in  $Q \supset A$ ) and whenever the antecedent is false (as is the case in  $X \supset Q$ ).

## EXERCISES

Compute the truth-values of the following compound sentences, where A, B, and C are true; X, Y, and Z are false; and P and Q are of unknown truth-value.

Sample Problem:      Sample Solution:

a)  $(A \vee X) \bullet (Z \supset C)$                       T

1.  $\sim B \vee X$
2.  $A \bullet \sim Z$
3.  $\sim X \supset \sim C$
4.  $(B \equiv C) \vee (X \bullet Y)$
5.  $\sim (C \vee (X \bullet \sim Y))$
6.  $(X \supset \sim A) \supset (\sim Z \bullet B)$
7.  $\sim (A \vee \sim X) \vee (C \bullet \sim Y)$
8.  $A \supset (\sim X \bullet \sim (C \equiv Y))$
9.  $\sim (\sim Z \supset \sim (\sim (A \equiv B) \vee \sim X))$
10.  $\sim (A \equiv \sim (\sim C \supset (B \bullet \sim X))) \equiv (\sim ((\sim Y \vee \sim A) \supset \sim B) \bullet (A \supset (((B \equiv \sim X) \equiv Z) \supset \sim Y)))$
11.  $\sim X \vee (Q \supset Z)$
12.  $\sim Q \bullet (A \supset (P \bullet \sim B))$
13.  $\sim (Q \bullet \sim Z) \supset (\sim P \supset C)$
14.  $\sim (\sim (A \bullet B) \supset ((P \vee \sim C) \equiv (\sim X \vee Q)))$
15.  $\sim P \vee (Q \vee P)$

## IV. Translating from English into SL

Soon we will learn how to evaluate arguments in SL—arguments whose premises and conclusions are SL sentences. In real life, though, we're not interested in evaluating arguments in some artificial language; we're interested in evaluating arguments presented in natural languages like English. So in order for our evaluative procedure of SL argument to

have any real-world significance, we need to show how SL arguments can be fair representations of natural-language counterparts. We need to show how to translate sentences in English into SL.

We already have some hints about how this is done. We know that simple English sentences are represented as capital letters in SL. We know that our operators—tilde, dot, wedge, horseshoe, and triple-bar—are the SL counterparts of the English locutions ‘not’, ‘and’, ‘or’, ‘if/then’, and ‘if and only if’, respectively. But there is significantly more to say on the topic of the relationship between English and SL. Our operators—alone or in combination—can capture a much larger portion of English than that short list of words and phrases.

## Tilde, Dot, and Wedge

Consider the word ‘but’. In English, it has a different meaning from the word ‘and’. When I say “Andrew Carnegie is rich and generous,” I communicate one thing; when I say “Carnegie is rich, but generous,” I communicate something slightly different. Both utterances convey the assertions that Carnegie is rich, on the one hand, and generous on the other. The ‘but’-sentence, though, conveys something more—namely, that there’s something surprising about the generosity in light of the richness, that there’s some tension between the two. But notice that each of those utterances is true under the same circumstances: when Carnegie is both rich and generous; the difference between ‘but’ and ‘and’ doesn’t affect the truth-conditions. Since the meanings of our SL operators are specified entirely in terms of their effects on truth-values, SL is blind to the difference in meaning between ‘and’ and ‘but’. Since the truth-conditions for compounds featuring the two words are the same—true just in case both components are true, and false otherwise—we can use the dot to represent both. ‘Carnegie is rich and generous’ and ‘Carnegie is rich, but generous’ would both be rendered in SL as something like ‘ $R \bullet G$ ’ (where ‘ $R$ ’ stands for the simple sentence ‘Carnegie is rich’ and ‘ $G$ ’ stands for ‘Carnegie is generous’). Again, switching from English into SL is a strategy for dealing with the messiness of natural language: to conduct the kind of rigorous logical analyses involved in evaluating deductive arguments, we need a simpler, tamer language; the slight difference in meaning between ‘and’ and ‘but’ is one of the wrinkles we need to iron out before we can proceed.

There are other words and phrases that have the same effect on truth-value as ‘and’, and which can therefore be represented with the dot: ‘although’, ‘however’, ‘moreover’, ‘in addition’, and so on. These can all be used to form conjunctions.

There are fewer ways of forming disjunctions in English. Almost always, these feature the word ‘or’, sometimes accompanied by ‘either’. Whenever we see ‘or’, we will translate it into SL as the wedge. As we discussed, the wedge captures the inclusive sense of ‘or’—one or the other, or both. The exclusive sense—one or the other, but not both—can also be rendered in SL, using a combination of symbols. ‘Salvador Allende or Radimiro Tomic will win the election, but not both’. How would we translate that into SL? Let ‘ $A$ ’ stand for ‘Allende will win’ and ‘ $T$ ’ stand for ‘Tomic will win’. We know how to deal with the ‘or’ part: ‘Allende will win or Tomic will win’ is just ‘ $A \vee T$ ’. How about the ‘not both’ part? That’s the claim, paraphrasing slightly, that it’s not the case that both Allende and Tomic will win; that is, it’s the negation of the conjunction: ‘ $\sim (A \bullet T)$ ’. So we have the ‘or’ part, and we have the ‘not both’ part; the only thing left is the word ‘but’ in between. We just learned how to deal with that! ‘But’ gets translated as a dot. So the proper SL translation of ‘Salvador Allende or Radimiro Tomic will win the election, but not both’ is this:

$$(A \vee T) \bullet \sim (A \bullet T)$$

Notice we had to enclose the disjunction, ‘ $A \vee T$ ’, in parentheses. This is to remove ambiguity: without the parentheses, we wouldn’t know whether the wedge or the (middle) dot was the main operator, and so the construction would not have been well-formed. In SL, the exclusive sense of ‘or’ is expressed with a conjunction: it conjoins the (inclusive) ‘or’ claim to the ‘not both’ claim— one or the other, but not both.

It is worth pausing to reflect on the symbolization of ‘not both’, and comparing it to a complementary locution—‘neither/nor’. We symbolize ‘not both’ in SL as a negated conjunction; ‘neither/nor’ is a negated disjunction. The sentence ‘Neither Tomic nor Beyoncé will win the election’ would be rendered as ‘ $\sim (T \vee B)$ ’; that is, it’s not the case that either Tomic or Beyoncé will win.

When we discussed the syntax of SL, it was useful to use an analogy to arithmetic to understand the interactions between tildes and parentheses. Taking that analogy too far in the case of negated conjunctions and disjunctions can lead us into error. The following is true in arithmetic:

$$-(2 + 5) = -2 + -5$$

We can distribute the minus-sign inside the parentheses (it's just multiplying by -1). The following, however, are *not* true in logic:<sup>5</sup>

$$\sim (p \bullet q) \equiv \sim p \bullet \sim q \quad [\text{WRONG}]$$

$$\sim (p \vee q) \equiv \sim p \vee \sim q \quad [\text{WRONG}]$$

The tilde cannot be distributed inside the parentheses in these cases. For each, the left- and righthand components have different meanings. To see why, we should think about some concrete examples. Let 'R' stand for 'Andrew Carnegie is rich' and 'G' stand for 'Andrew Carnegie is generous'. ' $\sim (R \bullet G)$ ' symbolizes the claim that Carnegie is not both rich and generous. Notice that this claim is compatible with his actually being rich, but not generous, and also with his being generous, but not rich. The claim is just that he's not *both*. Now consider the claim that ' $\sim R \bullet \sim G$ ' symbolizes. The main operator in that sentence is the dot; it's a conjunction. Conjunctions make a commitment to the truth of each of their conjuncts. The conjuncts in this case symbolize the sentences 'Carnegie is not rich' and 'Carnegie is not generous'. That is, this conjunction is committed to Carnegie's lacking both richness and generosity. That is a stronger claim than saying he's not both: if you say he's not both, that's compatible with him being one or the other; ' $\sim R \bullet \sim G$ ', on the other hand, insists that both are ruled out. So, generally speaking, a negated conjunction makes a different (weaker) claim than the conjunction of two negations.

There is also a difference between a negated disjunction and the disjunction of two negations. Consider ' $\sim (R \vee G)$ '. That symbolizes the sentence 'Carnegie is neither rich nor generous'. In other words, he lacks both richness and generosity. That's a much stronger claim than the one symbolized by ' $\sim R \vee \sim G$ '—the disjunction 'Either Carnegie isn't rich or he isn't generous'. He lacks one or the other quality (or both; the disjunction is inclusive). That's compatible with his actually being rich, but not generous; it's also compatible with his being generous, but not rich.

Did you notice what happened there? I used the same language to describe the claim symbolized by ' $\sim (R \bullet G)$ ' and ' $\sim R \vee \sim G$ '. Both merely assert that he isn't *both* rich and generous; he may be one or the other. I also described the claims made by ' $\sim (R \vee G)$ ' and ' $\sim R \bullet \sim G$ ' the same way. Both make the stronger claim that he lacks both characteristics. This is true in general: negated conjunctions are equivalent to the disjunction of two negations; and negated disjunctions are equivalent to the conjunction of two negations. The following logical equivalences are true:<sup>6</sup>

$$\sim (p \bullet q) \equiv \sim p \vee \sim q$$

$$\sim (p \vee q) \equiv \sim p \bullet \sim q$$

If you want to distribute that tilde inside the parentheses (or, alternatively, moving from right to left, pull the tilde outside), you have to change the wedge to a dot (and *vice versa*).

## Horseshoe and Triple-Bar

There are many English locutions that we can symbolize using the horseshoe and the triple-bar— especially the horseshoe. In fact, as we shall see, it's possible to render claims translated with the triple-bar using the horseshoe instead (along with a dot). We will look at a representative sample of the many ways in which conditionals and

5. The triple-bar is a logical equals-sign; it indicates that the components have the same truth-conditions (meaning).
6. They're often referred to as "DeMorgan's Theorems," after the nineteenth century English logician Augustus DeMorgan, who was apparently the first to formulate in the terms of the modern formal system developed by his fellow countryman and contemporary, George Boole. DeMorgan didn't discover these equivalences, however. They have been known to logicians since the ancient Greeks. We will return to DeMorgan's Theorems in Chapter 5.



biconditionals can be expressed in English, and talk about how to translate them into SL using the horseshoe and triple-bar.

The canonical presentation of a conditional uses the words ‘if’ and ‘then’, as in ‘If the sun explodes, then we will need lots of sunscreen’. But the word ‘then’ isn’t really necessary: ‘If the Sun explodes we will need lots of sunscreen’ makes the same assertion. It would also be symbolized as ‘ $E \supset N$ ’ (with ‘E’ and ‘N’ standing for the obvious simple components). The word ‘if’ can also be replaced. ‘Provided the sun explodes, a lot of sunscreen will be needed’ also makes the same claim. Things get tricky if we vary the placement of the ‘if’. Putting it in the middle of sentence, we get ‘Your pain will go away if you drink this herbal tea every day for a week’, for example. Compare that sentence to the one we considered earlier: ‘If you drink this herbal tea every day for a week, then your pain will go away’. Read one, then the other. They make the same claim, don’t they? Rule of thumb: whatever follows the word ‘if’, when ‘if’ occurs on its own (without the word ‘only’; see below), is the antecedent of the conditional. We would translate both of these sentences as something like ‘ $D \supset P$ ’ (where ‘D’ is for drinking the tea, and ‘P’ is for the pain going away).

The word ‘only’ changes things. Consider: ‘I will win the lottery only if I have a ticket’. A sensible claim, obviously true. I’m suggesting this is a conditional. Let ‘W’ stand for ‘I win the lottery’ and ‘T’ stand for ‘I have a ticket’. Which is the antecedent and which is the consequent? Which of these two symbolizations is correct:

$T \supset W$

or

$W \supset T$

To figure it out, let’s read them back into English as canonical ‘if/then’ claims. The first says, “If I have a ticket, then I’ll win the lottery.” Well, that’s optimistic! But clearly false—something only a fool would believe. That can’t be the correct way to symbolize our original, completely sensible claim that I will win only if I have a ticket. So it must be the second symbolization, which says that if I did win the lottery, then I had a ticket. That’s better. Generally speaking, the component occurring before ‘only if’ is the antecedent of a conditional, and the component occurring after is the consequent.

The claim in the last example can be put differently: having a ticket is a necessary condition for winning the lottery. We use the language of “necessary and sufficient conditions” all the time. We symbolize these locutions with the horseshoe. For example, being at least 16 years old is a necessary condition for having a driver’s license (in most states). Let ‘O’ stand for ‘I am at least 16 years old’ and ‘D’ stand for ‘I have a driver’s license’. ‘ $D \supset O$ ’ symbolizes the sentence claiming that O is necessary for D. The opposite won’t work: ‘ $O \supset D$ ’, if we read it back, says “If I’m at least 16 years old, then I have a driver’s license.” But that’s not true. Plenty of 16-year-olds don’t get a license. There are additional conditions besides age: passing the test, being physically able to drive, etc.

Another way of putting that point: being at least 16 years old is not a *sufficient* condition for having a driver’s license; it’s not enough on its own. An example of a sufficient condition: getting 100% on every test is a sufficient condition for getting an A in a class (supposing tests are the only evaluations). That is, if you get 100% on every test, then you’ll get an A. If ‘H’ stands for ‘I got 100% on all the tests’ and ‘A’ stands for ‘I got an A in the class’, then we would indicate that H is a sufficient condition for A in SL by writing ‘ $H \supset A$ ’. Notice that it’s not a necessary condition: you don’t have to be perfect to get an A. ‘ $A \supset H$ ’ would symbolize a falsehood.

To define a concept is to provide necessary and sufficient conditions for falling under it. For example, a bachelor is, by definition, an unmarried male. That is, being an unmarried male is necessary and sufficient for being a bachelor: you don’t qualify as a bachelor if you’re not an unmarried male, and being an unmarried male is enough, on its own, to qualify for bachelorhood. It’s for circumstances like this that we have the triple-bar. Recall, the phrase that triple-bar is meant to capture the meaning of is ‘if and only if’. We’re now in a position to understand that locution. Consider the claim that I am a bachelor if and only if I am an unmarried male. This is really a conjunction of two claims: I am a bachelor *if* I’m an unmarried male, *and* I’m a bachelor *only if* I’m an unmarried male. Let ‘B’ stand for ‘I’m a bachelor’ and ‘U’ stand for ‘I’m an unmarried male’. Our claim is then B if U, and B only if U. We know how to deal with ‘if’ on its own between two sentences: the one after the ‘if’ is the antecedent of the conditional. And we know how to deal with ‘only if’: the sentence before it is the antecedent, and the sentence after it is the consequent. To symbolize ‘I am a bachelor if and only if I am an unmarried male’ using horseshoes and a dot, we get this:

$$(U \supset B) \bullet (B \supset U)$$

The left-hand conjunct is the ‘if’ part; the right-hand conjunct is the ‘only if’ part. The purpose of the triple-bar is to give us a way of symbolizing such claims more easily, with a single symbol. ‘I am a bachelor if and only if I am an unmarried male’ can be translated into SL as ‘ $B \equiv L$ ’, which is just shorthand for the longer conjunction of conditionals above. And given that ‘necessary and sufficient’ is also just a conjunction of two conditionals, we use triple-bar for that locution as well. (Also, the phrase ‘just in case’ can be used to express a biconditional claim.)

At this point, you may have an objection: why include triple-bar in SL at all, if it’s dispensable in favor of a dot and a couple of horseshoes? Isn’t it superfluous? Well, yes and no. We could do without it, but having it makes certain translations easier. As a matter of fact, this is the case for all of our symbols. It’s always possible to replace them with combinations of others. Consider the horseshoe. It’s false when the antecedent is true and the consequent false, true otherwise. So really, it’s just a claim that it’s not the case that the antecedent is true and the conclusion false—a negated conjunction. We could replace any  $p \supset q$  with  $\sim (p \bullet \sim q)$ . And the equivalences we saw earlier—DeMorgan’s Theorems—show us how we can replace dots with wedges and *vice versa*. It’s a fact (I won’t prove it; take my word for it) that we could get by with only two symbols in our language: tilde and any one of wedge, dot, or horseshoe. So yeah, we have more symbols than we need, strictly speaking. But it’s convenient to have the number of symbols that we do, since they line up neatly with English locutions, making translation between English and SL much easier than it would be otherwise.

## EXERCISES

Translate the following into SL, using the bolded capital letters to stand for simple sentences.

Sample problem / Sample Solution

Everything will be **Allright** if we get to **Tir Asleen**. /  $T \supset A$

1. Harry Lime is a **Criminal**, but he’s not a **Monster**.
2. If **Thorwald** didn’t kill his wife, then **Jeffries** will look foolish.
3. Rosemary doesn’t love both **Max** and **Herman**.
4. Michael will not **Kill** Fredo if his **Mother** is still alive.
5. Neither **Woody** nor **Buzz** could defeat Zurg, but **Rex** could.
6. If either **Fredo** or **Sonny** takes over the family, it will be a **Disaster**.
7. **Eli** will get rich only if **Daniel** doesn’t drink his milkshake.
8. Writing a hit **Play** is necessary for Rosemary to fall in **Love** with Max.
9. Kane didn’t **Win** the election, but if the opening of the **Opera** goes well he’ll regain his **Dignity**.
10. If Dave flies into the **Monolith**, then he’ll have a **Transformative** experience; but if he doesn’t fly into the **Monolith**, he will be stuck on a **Ghost** ship.
11. Kane wants **Love** if and only if he gets it on his own **Terms**.
12. Either Henry keeps his **Mouth** shut and goes to **Jail** for a long time or he **Rats** on his friends and lives the rest of his life like a **Schnook**.
13. Only if Herman builds an **Aquarium** will Rosemary **Love** him.
14. **Killing** Morrie is sufficient for keeping him **Quiet**.
15. Jeffries will be **Vindicated**, provided Thorwald **Killed** his wife and Doyle **Admits** he was right all along.
16. **Collaborating** with Cecil B. DeMille is necessary to **Revive** Norma’s career, and if she does not **Collaborate** with DeMille, Norma may go **Insane**.
17. Either **Daniel** or **Eli** will get the oil, but not both.
18. To have a **Fulfilling** life as a toy, it is necessary, but not sufficient, to be **Played** with by children.
19. The Dude will get **Rich** if Walter’s **Plan** works, and if the Dude gets **Rich**, he’ll buy a new **Bowling** ball and a

new Carpet.

20. Either the AE-35 Unit is really malfunctioning or HAL has gone Crazy; and if HAL has gone Crazy, then the Mission will be a failure and neither Dave nor Frank will ever get home.

## V. Testing for Validity in SL

Having dealt with the task of taming natural language, we are finally in a position to complete the second and third steps of building a logic: defining logical form and developing a test for validity. The test will involve applying skills that we've already learned: setting up truth tables and computing the truth-values of compounds. First, we must define logical form in SL.

### Logical Argument Form in SL

This will seem trivial, but it is necessary. We're learning how to evaluate arguments expressed in SL. Like any evaluation of deductive arguments, the outcome hinges on the argument's form. So what is the form of an SL argument? Let's consider an example; here is an argument in SL:

$$\begin{array}{l} A \supset B \\ \sim B \\ \hline \therefore \sim A \end{array}$$

'A' and 'B' stand for simple sentences in English; we don't care which ones. We're working within SL: given an argument in this language, how do we determine its form? Quite simply, by systematically replacing capital letters with variables (lower-case letters like 'p', 'q', and 'r'). The form of that particular SL argument is this:

$$\begin{array}{l} p \supset q \\ \sim q \\ \hline \therefore \sim p \end{array}$$

The replacement of capital letters with lower-case variables was *systematic* in this sense: each occurrence of the same capital letter (e.g., 'A') was replaced with the same variable (e.g., 'p').

To generate the logical form of an SL argument, what we do is systematically replace SL sentences with what we'll call *sentence-forms*. An SL sentence is just a well-formed combination of SL symbols—capital letters, operators, and parentheses. A sentence-form is a combination of symbols that *would be* well-formed in SL, except that it has lower-case variables instead of capital letters.

Again, this may seem like a trivial change, but it is necessary. Remember, when we're testing an argument for validity, we're checking to see whether its form is such that it's *possible* for its premises to turn out true and its conclusion false. This means checking various ways of filling in the form with particular sentences. Variables—as the name suggests—can *vary* in the way we need: they are generic and can be replaced with any old particular sentence. Actual SL constructions feature capital letters, which are actual sentences having specific truth-values. It is conceptually incoherent to speak of checking different possibilities for actual sentences. So we must switch to sentence-forms.

### The Truth Table Test for Validity

To test an SL argument for validity, we identify its logical form, then create a truth table with columns for each of the variables and sentence-forms in the argument's form. Filling in columns of Ts and Fs under each of the operators in those sentence-forms will allow us to check for what we're looking for: an instance of the argument's form for which the premises turn out true and the conclusion turns out false. Finding such an instance demonstrates the argument's invalidity, while failing to find one demonstrates its validity.

To see how this works, it will be useful to work through an example. Consider the following argument in English:

If the Demigorgon returns, lots of lawns will be mowed.

The Demigorgon won't return.

$\therefore$  Lots of lawns will not be mowed.

We'll evaluate it by first translating it into SL. Let 'D' stand for 'the Demigorgon returns' and 'L' stand for 'Lots of lawns will be mowed'. This is the argument in SL:

$D \supset L$

$\sim D$

$\therefore \sim L$

First, the logical form. Replacing 'D' with 'p' and 'L' with 'q', we get:

$p \supset q$

$\sim p$

$\therefore \sim q$

Now we set up a truth table, with columns for each of the variables and columns for each of the sentence-forms. To determine how many rows our table needs, we note the number of different variables that occur in the argument-form (call that number 'n'); the table will need  $2^n$  rows. In this case, we have two variables—'p' and 'q'—and so we need  $2^2 = 4$  rows. (If there were three variables, we would need  $2^3 = 8$  rows; if there were four,  $2^4 = 16$ ; and so on.) Here is the table we need to fill in for this example:

variable	variable	premise 1	premise 2	$\therefore$
p	q	$p \supset q$	$\sim p$	$\sim q$

First, we fill in the "base columns". These are the columns for the variables. We do this systematically. Start with the right-most variable column (under 'q' in this case), and fill in Ts and Fs alternately: T, F, T, F, T, F, ... as many times as you need—until you've got a truth-value in every row. That gives us this:

p	q	$p \supset q$	$\sim p$	$\sim q$
	T			
	F			
	T			
	F			

Next, we move to the base column to the left of the one we just filled in (under 'p' now), and fill in Ts and Fs by alternating in twos: T, T, F, F, T, T, F, F, ... as many times as we need. The result is this:

p	q	$p \supset q$	$\sim p$	$\sim q$
T	T			
T	F			
F	T			
F	F			

If there were a third base column, we would fill in the Ts and Fs by alternating in fours: T, T, T, T, F, F, F, F.... For a fourth base column, we would alternate every other eight. And so on.

Next, we need to fill in columns of Ts and Fs under each of the operators in the statement-forms' columns. To do this, we apply our knowledge of how to compute the truth-values of compounds in terms of the values of their components, consulting the operators' truth table definitions. We know how to compute the values of  $p \supset q$ : it's false when p is true and q false; true otherwise. We know how to compute the values of  $\sim p$  and  $\sim q$ : those are just the opposites of the values of p and q in each of the rows. Making these computations, we fill the table in thus:

p	q	$p \supset q$	$\sim p$	$\sim q$
T	T	T	F	F
T	F	F	F	T
F	T	T	T	F
F	F	T	T	T

Once the table is filled in, we check to see if we have a valid or invalid form. *The mark of an invalid form is that it's possible for the premises to be true and the conclusion false.* Here, the rows of the table are the possibilities—the four possible outcomes of plugging in particular SL sentences for the variables: both true; the first is true, but the second false; the first false but the second true; both false. The reason we systematically fill in the base columns as described above is that the method ensures that our rows will collectively exhaust all these possible combinations.

So, to see if it's possible for the premises to come out true and the conclusion to come out false, we check each of the rows, looking for one in which this happens—one in which there's a T under ' $p \supset q$ ', a T under ' $\sim p$ ', and an F under ' $\sim q$ '. And we have one: in row 3, the premises come out true and the conclusion comes out false. This is enough to show that the argument is invalid:

v1	v2	prem. 1	prem. 2	$\therefore$
p	q	$p \supset q$	$\sim p$	$\sim q$
T	T	T	F	F
T	F	F	F	T
F	<b>T</b>	<b>T</b>	<b>T</b>	<b>F</b>
F	F	T	T	T

INVALID

When we're checking for validity, we're looking for one thing, and one thing only: a row (or rows) in which the premises come out true and the conclusion comes out false. If we find one, we have shown that the argument is invalid. If we set up and filled out the truth table correctly we have a row for every different possible scenario. And so, *if we do not find a row with true premises and a false conclusion, that indicates that it's impossible for the premises to be true and the conclusion false—and so the argument is valid.* Either way, the only thing we look for is a row with true premises and a false conclusion.

It only takes one row to invalidate an argument. Every other kind of row is irrelevant. It's common for beginners to mistakenly think they are. The fourth row in the table above, for example, looks significant. Everything comes out true in that row. Doesn't that mean something—something good, like that the argument's valid? No. Remember, each row only represents a *possibility*; what row 4 shows is that it's *possible* for the premises to be true and the conclusion true. But that's not enough for validity. For an argument to be valid, the premises must *guarantee* the conclusion; whenever they're true, the conclusion *must* be true. That it's merely *possible* that they all come out true is not enough.

Let's look at a more involved example, to see how the computation of the truth-values of the statement-forms must sometimes proceed in stages. The skill required here is nothing new—it's just identifying main operators and computing the values of the simplest components first—but it takes careful attention to keep everything straight. Consider this SL argument (never mind what its English counterpart is):

$$\begin{array}{l} (\sim A \bullet B) \vee \sim X \\ B \supset A \\ \hline \therefore \sim X \end{array}$$

To get its form, we replace 'A' with 'p', 'B' with 'q', and 'X' with 'r':

$$\begin{array}{l} (\sim p \bullet q) \vee \sim r \\ q \supset p \\ \hline \therefore \sim r \end{array}$$

So our truth-table will look like this (eight rows because we have three variables;  $2^3 = 8$ ):

p	q	r	$(\sim p \bullet q) \vee \sim r$	$q \supset p$	$\sim r$

Filling in the base columns as prescribed above—alternating every other one for the column under 'r', every two under 'q', and every four under 'p'—we get:

p	q	r	$(\sim p \bullet q) \vee \sim r$	$q \supset p$	$\sim r$
T	T	T			
T	T	F			
T	F	T			
T	F	F			
F	T	T			
F	T	F			
F	F	T			
F	F	F			

Now we turn our attention to the three sentence-forms. We'll start with the first premise, the compound ' $(\sim p \bullet q) \supset \sim r$ '. We need to compute the truth-value of this formula. We know how to do this, provided we have the truth-values of the simplest parts; we've solved problems like that already. The only difference in the case of truth tables is that there are multiple different assignments of truth-values to the simplest parts. In this case, there are eight different ways of assigning truth-values to 'p', 'q', and 'r'; those are represented by the eight different rows of the table. So we're solving a problem we know how to solve; we're just doing it eight times.

We start by identifying the main operator of the compound formula. In this case, it's the wedge: we have a disjunction; the left-hand disjunct is ' $(\sim p \bullet q)$ ', and the right-hand disjunct is ' $\sim r$ '. To figure out what happens under the wedge in our table, we must first figure out the values of these components. Both disjuncts are themselves compound: ' $(\sim p \bullet q)$ ' is a conjunction, and ' $\sim r$ ' is a negation. Let's tackle the conjunction first. To figure out what happens under the dot, we need to know the values of ' $\sim p$ ' and 'q'. We know the values of 'q'; that's one of the base columns. We must compute the value of ' $\sim p$ '. That's easy: in each row, the value of ' $\sim p$ ' will just be the opposite of the value of 'p'. We note the values under the tilde, the operator that generates them:

p	q	r	$(\sim p)$	$\bullet$	q)	$\vee$	$\sim r$	$q \supset p$	$\sim r$
T	T	T	F		T				
T	T	F	F		T				
T	F	T	F		F				
T	F	F	F		F				
F	T	T	T		T				
F	T	F	T		T				
F	F	T	T		F				
F	F	F	T		F				

To compute the value of the conjunction, we consider the result, in each row, of the truth-function for dot, where the inputs are the value under the tilde in ' $\sim p$ ' and the value under 'q' in the base column. In rows 1 and 2, it's  $F \bullet T$ ; in rows 3 and 4,  $F \bullet F$ ; and so on. The results:

p	q	r	( $\sim$ p	•	q)	$\vee$	$\sim$ r	$q \supset p$	$\sim$ r
T	T	T	F	F	T				
T	T	F	F	F	T				
T	F	T	F	F	F				
T	F	F	F	F	F				
F	T	T	T	T	T				
F	T	F	T	T	T				
F	F	T	T	F	F				
F	F	F	T	F	F				

The column we just produced, under the dot, gives us the range of truth-values for the left-hand disjunct in the first premise. We need the values of the right-hand disjunct. That's just ' $\sim$  r', which is easy to compute: it's just the opposite value of 'r' in every row:

p	q	r	( $\sim$ p	•	q)	$\vee$	$\sim$ r	$q \supset p$	$\sim$ r
T	T	T	F	F	T		F		
T	T	F	F	F	T		T		
T	F	T	F	F	F		F		
T	F	F	F	F	F		T		
F	T	T	T	T	T		F		
F	T	F	T	T	T		T		
F	F	T	T	F	F		F		
F	F	F	T	F	F		T		

Now we can finally determine the truth-values for the whole disjunction. We compute the value of the wedge's truth-function, where the inputs are the columns under the dot, on the one hand, and the tilde from ' $\sim$  r' on the other.  $F \vee F$ ,  $F \vee T$ ,  $F \vee F$ , and so on:

p	q	r	( $\sim$ p	•	q)	$\vee$	$\sim$ r	$q \supset p$	$\sim$ r
T	T	T	F	F	T	F	F		
T	T	F	F	F	T	T	T		
T	F	T	F	F	F	F	F		
T	F	F	F	F	F	T	T		
F	T	T	T	T	T	T	F		
F	T	F	T	T	T	T	T		
F	F	T	T	F	F	F	F		
F	F	F	T	F	F	T	T		

Since that column represents the range of possible values for the entire sentence-form, we highlight it. When we test for validity, we're looking for rows where the premises as a whole come out true; we'll be looking for the value under their main operators. To make that easier, just so we don't lose track of things visually because of all those columns, we highlight the one under the main operator.

Next, the second premise, which is thankfully much less complex. It is, however, slightly tricky. We need to compute the



value of a conditional here. But notice that things are a bit different than usual: the antecedent, 'q', has its base column to the *right* of the column for the consequent, 'p'. That's a bit awkward. We're used to computing conditionals from left-to-right; we'll have to mentally adjust to the fact that 'q  $\supset$  p' goes from right-to-left. (Alternatively, if it helps, you can simply reproduce the base columns underneath the variables in the 'q  $\supset$  p' column.) So in the first two rows, we compute T  $\supset$  T; but in rows 3 and 4, it's F  $\supset$  T; in rows 5 and 6, it's T  $\supset$  F (the only circumstance in which conditionals turn out false); and in rows 7 and 8, it's F  $\supset$  F. Here is the result:

p	q	r	( $\sim$ p	$\bullet$	q)	$\vee$	$\sim$ r	q $\supset$ p	$\sim$ r
T	T	T	F	F	T	F	F	T	
T	T	F	F	F	T	T	T	T	
T	F	T	F	F	F	F	F	T	
T	F	F	F	F	F	T	T	T	
F	T	T	T	T	T	T	F	F	
F	T	F	T	T	T	T	T	F	
F	F	T	T	F	F	F	F	T	
F	F	F	T	F	F	T	T	T	

No need to highlight that column, as it's the only one we produced for that premise, so there can be no confusion. We finish the table by computing the values for the conclusion, which is easy:

p	q	r	( $\sim$ p	$\bullet$	q)	$\vee$	$\sim$ r	q $\supset$ p	$\sim$ r
T	T	T	F	F	T	F	F	T	F
T	T	F	F	F	T	T	T	T	T
T	F	T	F	F	F	F	F	T	F
T	F	F	F	F	F	T	T	T	T
F	T	T	T	T	T	T	F	F	F
F	T	F	T	T	T	T	T	F	T
F	F	T	T	F	F	F	F	T	F
F	F	F	T	F	F	T	T	T	T

Is the argument valid? We look for a row with true premises and a false conclusion. There are none. The only two rows in which both premises come out true are the second and the eighth, and in both we also have a true conclusion. It is impossible for the premises to be true and the conclusion false, so the argument is valid.

So that is how we test arguments for validity in SL. It's a straightforward procedure; the main source of error is simple carelessness. Go step by step, keep careful track of what you're doing, and it should be easy. It's worth noting that the truth table test is what logicians call a "decision procedure": it's a rule-governed process (an algorithm) that is guaranteed to answer your question (in this case: valid or invalid?) in a finite number of steps. It is possible to program a computer to run the truth table test on arbitrarily long SL arguments. This is comforting, since once one gets more than four variables or so, the process becomes unwieldy.

## EXERCISES

Test the following arguments for validity using Truth Table method. For those that are invalid, specify the row(s) that demonstrate the invalidity.

For the problems, each premise is separated by a comma and the conclusion by our standard symbols. And so,

$$A \vee B, \sim A, / \therefore \sim B$$

can be thought of as:

$$A \vee B$$

$$\sim A$$

$$/ \therefore \sim B$$

1.  $\sim A, / \therefore \sim A \vee B$
2.  $A \supset B, \sim B, / \therefore \sim A$
3.  $A \vee B, \sim A, / \therefore \sim B$
4.  $\sim (A \equiv B), A \vee B, / \therefore A \bullet B$
5.  $A \supset B, A, / \therefore B$
6.  $A \supset B, B \supset C, / \therefore A \supset C$
7.  $\sim (A \supset B), \sim B \vee \sim A, / \therefore \sim (\sim A \equiv B)$
8.  $\sim B \vee A, \sim A, A \equiv B, / \therefore \sim B$
9.  $A \supset (B \bullet C), \sim B \vee \sim C, / \therefore \sim A$
10.  $\sim A \vee C, \sim B \supset \sim C, / \therefore A \equiv C$
11.  $\sim A \vee (\sim B \bullet C), \sim (C \vee B) \supset A, / \therefore \sim C \supset \sim B$
12.  $A \vee B, B \supset C, \sim C, / \therefore \sim A$

## Deductive Logic III: Natural Proofs in Propositional Logic

SEAN GOULD

### Deductive Logic III: Natural Proofs in Propositional Logic<sup>1</sup>

#### I. Replacement and Rules of Inference

In Chapter 4, we saw that we could use the Truth Table Test to check an argument's validity. If there is no possible scenarios where all true premises lead to a false conclusion then the argument is valid. This also means that a *valid argument* always preserves the potential truth of its premises all the way through to its conclusion. Therefore, it is also possible to ensure an argument's validity by ensuring that it can be established through a chain reasoning such that every link in the chain is valid itself. A major goal of this chapter is learning how to write out premises, conclusions, and the in-between steps in a very neat and tidy manner so that we can check each step and thus prove validity.

This idea of preserving truth introduces another way of establishing the validity of arguments, *natural deduction*. This is also called using “formal proofs of validity.” When we use natural deduction, we move through an argument one premise at a time, show what these premises legitimately entail, and eventually arrive at a statement of our conclusion. We reach our conclusion by taking certain baby steps that are guaranteed to preserve any potential truth of what came before them.

The two types of tools we use for basic formal proofs are the Rules of Replacement and the Rules of Inference. We use these tools as guarantors of individual steps within an overall valid chain of reasoning. And this is related to something many of us have done outside formal logic. When we argue and try to walk someone through our reasoning, we often try to explicitly show how the truth of one idea, or the combination of two ideas, entails the truth of a further idea. In doing so, we aim to express how some complicated, two-part statements (i.e., those involving multiple, singular propositions, such as “ $A \bullet H$ ” or “ $A \vee H$ ”) can further entail other claims when more information is given. We should be able to see that when given “ $A \supset B$ ” and “ $A$ ” we can conclude “ $B$ ”. A concrete example might be “If it's raining then the ground will be wet. It's raining. Thus, the ground is wet.” Or, if told, “the patient has a broken arm and a heart condition” we can legitimately conclude the simpler claim, “the patient has a heart condition,” from the earlier compound statement. Formally, this would be like concluding “ $H$ ” from “ $A \bullet H$ .” And it might just be that this simpler *intermediary premise* is just what we'd need to use in a proof of some other conclusion, such as that the patient either has a heart condition or Lyme disease, “ $H \vee L$ .”

The above reasoning informally utilized some rules of inference. The first section of this chapter will cover nine such *rules of inference* and give them labels for future use as tools in constructing formal proofs. The second section covers the rules of replacement and shows us how some statements in SL are equivalent to each other. For example, we've already seen that “ $\sim\sim p$ ” is equal to “ $p$ .” We also saw that “ $\sim(p \bullet q)$ ” is true if and only if either  $p$  or  $q$  is not true, i.e., “ $\sim p \vee \sim q$ .” We will see that when we talk about replacement rules, we are just looking at statements that are really saying the same thing in different ways. This means that whenever the one is true whenever the other is, and vice versa, and so

1. This chapter follows Copi, I.M., Cohen, C. & Rodych, V. (2020) Introduction to Logic. 15th Ed. Special Indian Edition. Routledge. Chapter 9, pp 332 - 450. for its organization and general content presentation. Where direct excerpts are utilized, specific citation is given through further footnotes.

logically speaking, the two are functionally equivalent. However, restating the expression of the term, such as replacing “ $\sim\sim p$ ” with “ $p$ ,” can often help us move closer to expressing a particular conclusion. Therefore, the second part of this chapter will present 10 such *replacement rules* and give them labels for use as tools in constructing formal proofs.

These inference and replacement rules are not exhaustive of all valid arguments steps. In fact, we will introduce a few more later in the book. However, these 19 rules of inference and replacement are enough to help justify almost any basic deductive argument. These “rules” make up a set of reasoning steps we can rely upon to ensure or prove valid reasoning.

After we’ve got a labeled set of 19 rules of inference and replacement rules, the third section will explain how we write out a formal proof using these new tools. As a quick preview of where we are going, our aim will be to construct little roadmaps where the premises and their validly-entailed intermediate premises are written in a “statement column” on the left, and the justification for adding them (one of the 19 rules of inference or replacement rules) is provided in a “justification ledger” on the right. After the third section of this chapter, our aim is to be able to understand the following notation for proving  $D$  from  $A \bullet B$  and  $(B \vee C) \supset D$  would look like this:

---

Formal Proof Example 1

1.	$A \bullet B$	
2.	$(C \vee B) \supset D$	$\therefore D$
3.	$B$	1, Simp.
4.	$B \vee C$	3, Add.
5.	$C \vee B$	4, Com.
6.	$D$	1, 2, M.P.

---

It’s okay if this looks cryptic now. The above method of writing out proofs will all get explained in the following sections. Our learning goal is to work towards understanding this and being able to write things out in this manner. By the end of this chapter, you will be able to see how the first two lines are our given premises. You will be able to recognize that the right column states our end conclusion as  $\therefore D$ , and that line 3 uses the rule “simplification” to get  $B$  from the premise  $A \bullet B$  that was stated in line 1, as is noted in the right-hand column. You will also be able to understand how line 4 uses a rule called “addition” to move from line 3 to line 4, and at the end of this chapter, you will understand how line 5 uses a rule called *modus ponens* to get  $D$  from lines 2 and 4.

Writing out deductive arguments in this systematic way allows us to focus our attention on the form of the argument alone. We can avoid getting distracted by details, idiosyncratic ways of writing terms, or other tangential issues. By training ourselves to utilize a consistent presentation of arguments, we can make analyzing and talking about arguments a little easier. Learning this skill may seem hard at first, but it will save effort once it is mastered, sort of like how learning to ride a bike is more difficult than walking, at first. Once we’ve got things figured out the bike becomes a very useful tool.

After we introduce the basic idea of proofs and how to read and write them, the chapter will finish by adding two more argumentative steps. The first is using hypothetical scenarios alongside given premises to establish certain conditional connections which we can further use in our arguments – so-called “conditional proofs.” The other is how to formally show potential contradictions between propositions to establish statements of negation, so-called “indirect proofs.”

A person doesn’t need to memorize all 19 of these tools and the two tricks to do natural deduction and write out formal proofs. Most of us create little proofs all the time anyhow. We’re just learning how to articulate them so that the connections are made more salient, transparent, and explicit. And so, there’s no reason a cheat sheet of the 19 rules wouldn’t do just as well as memorizing them. However, there is a bit of creativity involved in finding an elegant way to

validly move from premises to a conclusion while showing every step. Having the 19 tools we are about to encounter at one's fingertips helps us know which ones to use and when.

## II. Rules of Inference

### *Modus Ponens (M.P.)*

*Modus ponens* (M.P. for short) is the first “rule of inference” we will cover. The reasoning itself is very, very basic and might be familiar to you. What we are doing, overall, though, is aiming to recognize some very basic patterns across multiple examples and talk about them in a way that will serve other purposes, such as expressing formal proofs later down the road.

And so, our first “rule of inference” utilizes a familiar argument form. Suppose we claim “p” and “if p, then q.” Well, then, we’ve set ourselves up for asserting “q.” This simple inference is called “*modus ponens*.”

Notice how the “example” we just used was very schematic. That’s because it was only the *general form*. An argument form shows the relationships between potential propositions without relying upon any specific content. It’s even more abstract than using capital letters to stand in for specific propositions.

Let’s step back a step to review the notion of *argument form* (sometimes called “logical form, or “general form”). In previous chapters, we talked about how we can take full sentences from English, Spanish, or any language and boil them down to their basic propositions. We also saw that it is quite convenient to take particular propositions and label them with a capital letter. “Snow is White” becomes “W” – if only so that we needn’t write the whole thing down.

Now, suppose we say, “if the snow is white, then the snow is clean,” and we want to use W for “the snow is white,” and C for “the snow is clean,” and “ $\supset$ ” for the “if . . . then . . .” relationship. Let’s further say we’ve got an argument with the premise “if the snow is white, then the snow is clean” written out as “ $W \supset C$ ,” and we also add the premise “the snow is white” written as just plain “W.” And suppose from these two we conclude “the snow is clean.” We will add a new symbol, “ $\therefore$ ” to stand for “in conclusion.”

And so, we’ve got the argument:

---

$$\begin{array}{l} W \supset C \\ W \\ \therefore C \end{array}$$

---

Okay. Now, suppose the next day we give a different argument: “if the dog is howling, then the cat will be annoyed. And the dog is howling. Therefore, the cat is annoyed.” Say we want to use H for “the dog is howling,” and A for “the cat will be annoyed” and all the other same logical operator symbols as before.

And so, we’ve got the argument:

---

$$\begin{array}{l} H \supset A \\ H \\ \therefore A \end{array}$$

---

If we put the two arguments side by side, you should be able to see that even though the propositions differ (one's about snow, the other pets) the *argument form* is the exact same.

---

M. P. Example 1	M.P. Example 2
$H \supset A$	$W \supset C$
$H$	$W$
$\therefore A$	$\therefore C$

---

Now, in Chapter 3, III, “Semantics of SL” we encountered “argument forms” as a way to say things about our logical arguments. Here, to say something about the structure or “form” of *any given argument*, we’ll use lower case symbols such as  $p$ ,  $q$ ,  $r$ ,  $s$ , etc. as *variables* to stand in as functional placeholders for propositions (or their capital letter abbreviations).

---

M.P. Example 1	M.P. Example 2	M.P. Argument Form
$H \supset A$	$W \supset C$	$p \supset q$
$H$	$W$	$p$
$\therefore A$	$\therefore C$	$\therefore q$

---

Both examples use the same *argument form*. Any set of specific arguments that have the same general form is said to be “logically analogous.” In this case, the general form is called “*modus ponens*.” Here it is presented below, with its name, an abbreviation (useful for later formal-proof-making), and a reminder that it is always valid:

---

<i>Modus Ponens</i> (M.P.)
$p \supset q$
$p$
$\therefore q$
Always Valid!

---

The general form of *modus ponens* is always valid. If a specific argument has this form, such as in examples 1 and 2, then it is valid. Arguments with this form might not always be sound, but it is impossible for the two premises to be true and the conclusion false. For proof, you can refer to truth table proof exercise # 5 at the end of the last chapter. An accurate truth table will show that it is impossible for the two premises to be true and the conclusion false. The form says that if you have a conditional and affirm the antecedent, then the consequent follows as being true, too.

Once we accept a general form as valid, the application of the inference rule allows us to work backward to a specific argument. Thus, we can say something like

$$A \supset B$$

$$\begin{array}{c} A \\ \therefore B \end{array}$$

is justified by *modus ponens*.

The above argument instantiates a version of *modus ponens* and is therefore valid. Using rules of inference and replacement rules to justify an argument's validity is a critical skill in general logic and the construction of formal proofs.

Do not confuse *modus ponens* with a similar, pernicious, and fallacious line of reasoning called "affirming the consequent." Affirming the consequent is like trying to do *modus ponens* backward! Here is this mistake's general form:

---

Affirming the Consequent

$$\begin{array}{l} p \supset q \\ q \\ \therefore p \end{array}$$

Never Valid!

---



Notice how we switched the “q” and “p” in the bottom two lines of the form. This argument form does not preserve truth. It is possible for the premises – whatever gets plugged in for “ $p \supset q$ ” and “q” to be true while the antecedent fails to hold. Here’s an example to help illustrate such a scenario. Suppose we say, “if aliens built the pyramids, then the pyramids would have unsolved mysterious about them (that’s our content for “ $p \supset q$ ”). The pyramids do have unsolved mysteries (here’s the “q” part). Therefore, aliens built them (p).” Nope. There are many ways for mysteries to occur beyond extraterrestrial construction. Remember from Chapter 3 that a conditional can be true even when the antecedent does not obtain.

Any argument that partakes in this invalid form is logically analogous to any other argument that shares this same nasty form. Sometimes though, it can be hard to see that there is a problem.

“Affirming the consequent” is often a seductive fallacy; it is common in a lot of pseudosciences, wishful thinking, superstition, conspiracy theory, and other practices that utilize confirmation bias and the cherry-picking of evidence. However, as the above example shows, it is not a valid form of reasoning.

Sometimes it’s easier to show the invalidity of an argument by providing a logical analogy that is clearly off the mark. When we can see an example where the premise can be true while the conclusion of the argument can be false, we have reason to reject *any* argument with the same form. The fact that “if a person is eaten by a shark then they will be dead” is not enough to let us conclude that all dead people died from shark attack! On the other hand, valid argument forms, like *modus ponens*, force a true conclusion from true premises. No counterexamples can be drawn that match the premises but not the conclusion. This is what the truth-table process establishes.

There are many different potential general forms for arguments besides *modus ponens*. Being familiar with some of the most basic valid forms allows us to use them as rules of inference when we move on to further developing proofs. And so, let’s look at some more.

## Modus Tollens (M.T.)

Our second rule of inference is called *modus tollens* (abbreviated as M.T.). *Modus tollens* also uses a conditional, but it differs from *modus ponens*. Instead of establishing the consequent by asserting the antecedent, *modus ponens* works in the opposite direction. Here, we assert the conditional, deny the conclusion, and therefore commit ourselves to accept that the antecedent doesn’t hold either. For example, “If it’s nighttime in Boise, then it’s nighttime in Nampa. But it’s not nighttime in Nampa. Therefore, it’s not nighttime in Boise, either.” Here’s another example, “If Fred is guilty, then George must be guilty. But George is not guilty. And so, Fred must not be guilty after all.” Both tiny arguments utilize the form of *modus tollens*.

Let’s express the argument form occurring here. Label the propositions in the Boise/Nampa example as B = it’s nighttime in Boise, N = it’s nighttime in Nampa.  $\sim N$  = not nighttime in Nampa, and  $\sim B$  = it’s not nighttime in Boise. And for the Fred/George Example, let’s use F for Fred is guilty,  $\sim F$  for Fred’s not guilty, and G and  $\sim G$  for the same things for George. This gives us the following table with the two example arguments and their general form, all displayed side by side:

Example 1	Example 2	Argument Form
$B \supset N$ $\sim N$ $\therefore \sim B$	$F \supset G$ $\sim G$ $\therefore \sim F$	$p \supset q$ $\sim q$ $\therefore \sim p$

The general form here is called *modus tollens* (M. T.). Here it is by itself:

---

*Modus Tollens (M. T.)*

$$\begin{array}{l} p \supset q \\ \sim q \\ \therefore \sim p \end{array}$$

Always Valid!

---

*Modus tollens* is always valid. It is impossible for the premises to be true and the conclusion false. This, too, is displayed by the way the truth table (exercise # 2) maps out the possibilities. Conversely, we can justify the reasoning of any argument that instantiates this general form by simply saying it is, in fact, a version of *modus tollens*.

Do not confuse *modus tollens* with a similar, fallacious line of reasoning called “denying the antecedent.” Here is this mistake’s general form:

---

*Denying Antecedent*

$$\begin{array}{l} p \supset q \\ \sim p \\ \therefore \sim q \end{array}$$

Never Valid!

---

This argument form does not preserve truth. It is possible for the premises that fill in the “ $p \supset q$ ” and “ $\sim p$ ” parts to be true while the conclusion “ $\sim q$ ” is false. Here’s an example to help illustrate such a scenario. Suppose we say, “if it were to rain, then the grass would get wet (that’s our “ $p \supset q$ ” part). But it did not rain (here’s the  $\sim p$  part). Therefore, the grass will not be wet ( $\sim q$ ).” Lack of rain doesn’t guarantee dry lawns. Maybe sprinklers came on overnight, for example. The conditional in the first line, “if it were to rain, then the grass would get wet,” is true even if it doesn’t end up raining. Remember from Chapter 3 that a conditional can be true even when the antecedent does not in fact obtain.

### *Hypothetical Syllogism (H.S.)*

Hypothetical Syllogism refers to an argument form that links two conditionals into a longer chain. For example, “if I knock over the first domino, then the second will fall (premise 1). And, if the second one falls, the third will fall (premise 2). Therefore, if I knock over the first domino, the third will fall (conclusion).” Our speaker here seems pretty sure of her domino stacking, but if we take her word about the certainty of the premises (i.e., assume them as true) then we’d have to admit that the conclusion must hold. Thus, hypothetical syllogism represents our third valid argument form. You can see that all true premises yield a true conclusion if you refer to the correct answer for exercise # 6 from the end of the last chapter.

Here is hypothetical syllogism laid out in its argument form:

---

### Hypothetical Syllogism (H.S.)

$$\begin{array}{l} p \supset q \\ q \supset r \\ \therefore p \supset r \end{array}$$

Always Valid!

---

Just like dominos, right? Notice here that the conclusion is only asserting that the conditional chain can be linked. It does not assert that the final consequent obtains. For that, we'd need an added premise of one of the antecedents occurring. Returning to our domino example, we can deduce that knocking the first domino *would* cause the third one to fall without actually knocking any. If we take the first two conditionals as certain (we really trust the domino-set up), then we're committed to the conclusion. Denying a premise, i.e., saying a premise is false, and thereby questioning the *soundness* of the argument is not the same as rejecting the validity of the argument form itself. With *hypothetical syllogism*, a true conclusion always follows from true premises.

### *Disjunctive Syllogism (D.S.)*

Our fourth valid form is called “disjunctive syllogism.” You might already be familiar with it under another name, “the process of elimination,” where we are presented with two potential options, reject one, and therefore find ourselves left with the last standing alternative. For a micro-example, consider a coin toss: “The coin landed head or tails. It didn't land heads. Therefore, it landed tails.” Here's the argument form:

---

### Disjunctive Syllogism (D.S.)

$$\begin{array}{l} p \vee q \\ \neg p \\ \therefore q \end{array}$$

Always Valid!

---

If you did exercise # 3 from the last chapter correctly, you've already proved the form's validity.

A thing to remember here is that the argument form is taking the initial “ $p \vee q$ ” disjunct as true. Again, contesting a premise isn't the same as challenging the logical form. Keep *soundness* and *validity* as separate concepts. It is true that many times we can get tricked by a “false dichotomy” where we accidentally take a disjunct as true when it isn't. But that doesn't mean there are many legitimate disjuncts out there. Either way, and once again, whenever the premises *are* in fact true, with this valid argument form, the conclusion will hold.

### *Constructive Dilemma (C.D.)*

Suppose you are standing in front of two doors, and you get to open one, the other, or both. Door #1 leads to a lion. Door #2 leads to a lamb. Well, it seems safe to phrase your options as being the lion or the lamb (or both).

The form of reasoning we just used is called “constructive dilemma.” We took the dilemma of the two doors and built it out a little further – constructed – it to a different dilemma, lion  $\vee$  lamb.

Here's the form:

---

Constructive Dilemma (C.D.)

$$\begin{array}{l} (p \supset q) \bullet (r \supset s) \\ p \vee r \\ \therefore q \vee s \end{array}$$

Always Valid!

---

Keeping our example in mind, you can see how in premise #2, “ $p \vee r$ ” could represent the two doors. Premise #1,  $(p \supset q) \bullet (r \supset s)$ , gives a diagram of where the doors lead. And the conclusion “ $q \vee s$ ” just states where we could end up.

### *Simplification (Simp.)*

Let’s introduce our next form, “simplification,” with an example: “I went to the post office and the bank. So, yes, I went to the post office.” That’s it. Simplification just lets us recognize that when we assert a pair of claims in a conjunction, we are entitled to assert a simplified part that pair. Here’s the form:

---

Simplification (Simp.)

$$\begin{array}{l} p \bullet q \\ \therefore p \end{array}$$

Always Valid!

---

Sometimes formal logic can seem tricky simply because it’s hard to remember how fine-grained and simple the baby steps we are taking really are. But these small, subtle steps are useful. Sometimes isolating an important claim from a long string of statements really helps our reasoning move along.

### *Conjunction (Conj.)*

If we state two things separately, there’s no harm in stating them in one breath. Consider, “I have apples. I have oranges. Therefore, I have apples and oranges.” The argument form of “conjunction” might strike you as being so rudimentary that it’d be easy to overlook. Nevertheless, it is representing a way to move from two premises to a conclusion that always necessarily follows. Plus, keeping in mind that “conjoining” premises is a step we can take that will help us think clearly about more complex arguments down the road. Here’s the conjunction’s argument form:

---

Conjunction (Conj.)

$$\begin{array}{c} p \\ q \\ \therefore p \bullet q \end{array}$$

Always Valid!

---

### *Absorption (Abs.)*

When we assert a basic conditional ( $p \supset q$ ), we are saying that the consequent follows from the antecedent. This next argument form shows us that when we assert a conditional, then we also get to assert that both the antecedent and the consequent follow from the antecedent! Here's an example, "if the toast is burnt, then it will be black. Therefore, if the toast is burnt, it will be black and burnt." It may be funny to say that if the antecedent is true, then in addition to the consequent, the antecedent is true too, but it sure is valid. We are allowed to "absorb" the antecedent back into the consequent of its own conditional.

---

Absorption (Abs.)

$$\begin{array}{c} p \supset q \\ p \supset (p \bullet q) \end{array}$$

Always Valid!

---

### *Addition (Add.)*

Remember that we call things with the form ( $p \vee q$ ) a disjunction. And that the minimum truth requirement for a disjunction to be true is that at least one of its disjuncts holds. This means that when we assert a simple proposition, "p," then we could add on an "or q" for free!

Here's an example. "This is the last rule of inference presented in this section. Therefore, either this is the last rule of inference presented in this section, or the moon is made of cheese." This tiny argument, like all others of the same form, can be mapped out as such:

---

Addition (Add.)

$$\begin{array}{c} p \\ \therefore p \vee q \end{array}$$

Always Valid!

---

When the premise, in this case merely “p” holds, a disjunctive conclusion that includes the known part will be true, too. An accurate truth table for exercise # 1 from the end of last chapter confirms the validity of this simple argument form.

## Conclusion

The argument forms introduced in this section represent valid logical connections between premises and conclusions at a very fundamental level. In real life, arguments are often much more complicated. However, as we move on to constructing formal proofs of deductive arguments, it is important that we learn to recognize these forms so that we can appeal to them during each step of more complex chains of reasoning. This lets us assess arguments by breaking them into smaller pieces.

We’ve identified argument forms where given a true premise or two the conclusion must follow. In other words, we might say for our valid argument forms, “if the premises were true, then the conclusion would be true.”

Valid rules of inference DO NOT give us premises and conclusions that are interchangeable. It would be an instance of “affirming the consequent” to move from the conclusion of a valid argument form to an affirmation of its premises.

Simplification		Reverse?	
$p \bullet q$	And so?	$p$	
$\therefore p$	?	$\therefore p \bullet q$	
Valid Form		Reverse Doesn't Work	

While it is valid to say, “A • B, therefore A,” it is not valid to say, “A just means ‘A • B.’” To make it more concrete, I can’t say, “I have apples. Therefore, and from that fact alone, I have oranges, too.” This would be nonsense.

The next section on Replacement Rules will be different. In that section, we introduce propositional statements that are interchangeable because they are logically equivalent. “I don’t not have apples” really does mean the same as “I have apples.” The section on Replacement Rules will introduce accurate ways of re-stating claims – a skill that, in combination with our use of inference rules, will help us develop proofs of more complicated arguments.

## III. Replacement Rules

Before moving on to our replacement rules, it is important that we are comfortable with understanding just what our logical operators  $\sim$ ,  $\bullet$ ,  $\vee$ ,  $\supset$ , and  $\equiv$  mean. I.e., what they are saying – what their truth conditions are. Please refer to Chapter 3, Section III if you need to refresh your grasp. Understanding replacement rules simply builds from grasping the meaning of logical operators so that truth is maintained through different expressions of complex propositions. Especially, refer back to Chapter 3, III, “Biconditional (Triple-Bar)”  $\equiv$  which is introduced as logic’s “equals-sign.” When we have “a statement form all of whose substitution instances must be true” the equivalence is known as a “tautology” (Copit et. al). Tautologies are written as  $\equiv^t$  to show that the biconditional is of this special variety.

At the end of this section, you should be able to look at a general form of any given replacement rule and see how either way of stating a claim is equivalent to its sister claim. The ability to recognize this similarity can then help take any given claim and translated it to its sister expression when we move on to applying these replacement rules in the next section on formal proofs.

## Double Negation (D.N.)

You are probably familiar with the concept of a “double negative.” For any given proposition, negating the negation is the same as stating the proposition. For example, take “it’s not, not snowing.” We’ll use two negations, one for each “not,” and let the proposition that it is snowing = S so that our example can be written in SL as, “ $\sim\sim S$ .” Well, I hope you can see that  $\sim\sim S$  is true if and only if S is true.

In other terms, “ $\sim\sim S \equiv^t S$ .”

Of course, this feature of double negation holds for any proposition, not just S. So, we’ll go back to using the lower case variable “p” to express the form of the idea:

---

Double Negation (D.N.)

$$\sim\sim p \equiv^t p$$

---

Because each side of the “ $\equiv^t$ ” mean the exact same thing, the two can be interchanged or used to replace one another. Hence, *double negation* is our first replacement rule.

## Commutation (Com.)

Both the “ $\vee$ ” disjunction and the “ $\bullet$ ” conjunction signs present pairs of simple propositions. The truth of the complex disjunction ( $p \vee q$ ) or complex conjunction ( $p \bullet q$ ) do not depend on the order in which the components “p” and “q” are written. Therefore, we can use the “commutation” replacement rule to switch the order of the parts. Here is the form for commutation:

---

Commutation (Com.)

$$\begin{array}{l} (p \vee q) \equiv^t (q \vee p) \\ (p \bullet q) \equiv^t (q \bullet p) \end{array}$$

---

This form states that we can swap the order of propositions that are within a *disjunction* or *conjunction*. Because we want to show how every single change is justified by a valid step, even something as small as altering the order of components of a disjunction or a conjunction requires explicit justification. It is important to note, though, that just like in Ghost Busters, you cannot cross the streams. Each line only works with itself, and you cannot recombine the biconditionals. “ $p \vee q$ ” is not equivalent to “ $q \bullet p$ .”

## Association (Assoc.)

The replacement rule of *association* is like commutation. Association refers the fact that when we have a set of three disjuncts, or a set of three conjuncts, it is immaterial as to where we put the dividing parenthesis.

---

Association (Assoc.)

$$\begin{aligned} [p \vee (q \vee r)] &\equiv^t [(p \vee q) \vee r] \\ [p \bullet (q \bullet r)] &\equiv^t [(p \bullet q) \bullet r] \end{aligned}$$

---

Once again, the lines cannot cross.

### *Transposition (Trans.)*

Recall our friend *modus tollens*. That argument form said that if we have a conditional ( $p \supset q$ ) and reject the consequent “ $\sim q$ ,” it is valid to reject the antecedent and get “ $\sim p$ .” To rephrase this, if we know ( $p \supset q$ ) we know ( $\sim q \supset \sim p$ ). And these two always go together. We only have one if and only if we have the other. They are logically equivalent. Hence, we have another replacement rule, that of transposition:

---

Transposition (Trans.)

$$p \supset q \equiv^t (\sim q \supset \sim p)$$

---

### *Material Implication (Impl.)*

The replacement rule “material implication” follows from the definition of the horseshoe conditional sign. Recall the only way a conditional statement  $p \supset q$  is false is if “ $p$ ” is true, but “ $q$ ” is false. This means that if either  $p$  is false, or  $q$  is true, i.e.,  $(\sim p \vee q)$ , then the conditional  $p \supset q$  is true. And if the conditional  $p \supset q$  is true only if  $(\sim p \vee q)$  is true. Therefore:

---

Material Implication (Impl.)

$$p \supset q \equiv^t (\sim p \vee q)$$

---

### *De Morgan's Theorems (De M.)*

We met De Morgan's Theorems in Chapter 3. De Morgan's Theorems state what needs to be the case to negate a conjunction or a disjunction. A conjunction,  $p \bullet q$ , can be negated if and only if at least one of its conjuncts is negated. Hence,  $\sim(p \bullet q) \equiv^t (\sim p \vee \sim q)$ .

On the other hand, since a disjunct only needs one of its parts to hold for the entire disjunction to hold, a disjunction,  $p \vee q$ , can be negated if and only if both disjuncts are negated. And so,  $\sim(p \vee q) \equiv^t (\sim p \bullet \sim q)$ .

Therefore, we have the pair:



---

De Morgan's Theorems (De M.)

$$\begin{aligned}\sim(p \bullet q) &\equiv^t (\sim p \vee \sim q) \\ \sim(p \vee q) &\equiv^t (\sim p \bullet \sim q)\end{aligned}$$


---

Again, you cannot mix and match these two different forms of De Morgan's.

### *Distribution (Dist.)*

Suppose we know “roses are red” (R) and either “violets are blue (B) or violets are orange (O).” We could write this out in SL as  $R \bullet (B \vee O)$ . But this means, in other terms, that we know either “roses are red and violets are blue” or “roses are red and violets are orange.” Which we could write as  $(R \bullet B) \vee (R \bullet O)$ .

Hence, we see  $[R \bullet (B \vee O)] \equiv^t [(R \bullet B) \vee (R \bullet O)]$ . Forgetting flowers, we can articulate this logical equivalence schematically as:

$$[p \bullet (q \vee r)] \equiv^t [(p \bullet q) \vee (p \bullet r)]$$

The above gives us the first half of the *distribution* replacement rule. Here, the single conjunct is “distributed” to the other parts of the disjunctive to which it is conjoined.

There is a related variation involving an initial disjunction. Suppose we know “Al Capone robbed the bank (A) or both Selman robbed the bank (S), and Louise robbed the bank (L).” We can write it out as  $A \vee (S \bullet L)$ . Again, this same complex statement is logically equivalent to affirming both it was Al Capone or Selma, and it was Al Capone or Louise;  $(A \vee S) \bullet (A \vee L)$ . And so, we have  $A \vee (S \bullet L) \equiv (A \vee S) \bullet (A \vee L)$ . Since this is true of any set of propositions, we'll just write it schematically as:

$$[p \vee (q \bullet r)] \equiv^t [(p \vee q) \bullet (p \vee r)]$$

Thus, we have the complete statement of the *distribution* replacement rule:

---

Distribution (Dist.)

$$\begin{aligned}[p \bullet (q \vee r)] &\equiv^t [(p \bullet q) \vee (p \bullet r)] \\ [p \vee (q \bullet r)] &\equiv^t [(p \vee q) \bullet (p \vee r)]\end{aligned}$$


---

Again, each line only works with itself. You can't mix and match the pairs of the different biconditionals.

### *Material Equivalence (Equiv.)*

The *material equivalence* (Equiv.) replacement rule works by the nature of the biconditional itself. It also comes in two forms. First, to say  $p \supset q$  is the same as saying if p then q and if q then p. The second form states that if we say  $p \supset q$  then we are saying either you have both p and q, or you have neither. Schematically, we get:

---

Material Equivalence (Equiv.)

$$\begin{aligned}(p \equiv q) &\equiv^t [(p \supset q) \bullet (q \supset p)] \\ (p \equiv q) &\equiv^t [(p \bullet q) \vee (\sim p \bullet \sim q)]\end{aligned}$$


---

## *Exportation (Exp.)*

*Exportation* describes the fact that stating that pair of conjuncts entails a third term is logically equivalent to stating that if we know the first half of the pair, then we know that completing the pair would also yield the third term.

Here is an example. This first statement, “if the bucket is full (B) and it tips over (T) then the floor will get wet (W)” is logically equivalent to “if the bucket is full, then if it tips over the floor will get wet.” In our example, we have:  $[(B \bullet T) \supset W] \equiv [B \supset (T \supset W)]$ . Once again, replacing our proposition symbols with some general placeholders, we can express the form of this logical equivalence as such:

---

Exportation (Exp.)

$$[(p \bullet q) \supset r] \equiv^t [p \supset (q \supset r)]$$

---

## *Tautology (Taut.)*

The final replacement rule simply states that any proposition is logically equivalent to a conjunction of itself with itself or a disjunction of itself with itself. It is a tautological equivalence that is, itself, called “tautology.” In other words,  $p \equiv^t p \bullet p$  as well as  $p \equiv^t p \vee p$ . Stating this fact as a replacement rule gives a “formal stamp of approval” for replacing somewhat redundant statements with a simpler version. Sometimes during formal proofs, the outcome of a deduction, such as a constructive dilemma, might yield something like  $p \vee p$  from which we’d like to just conclude  $p$ . With this last rule, we can make all our steps clear and transparent.

Here is the schema for this final replacement rule:

---

Tautology (Taut.)

$$\begin{array}{l} p \equiv^t (p \vee p) \\ p \equiv^t (p \bullet p) \end{array}$$

---

We have now been introduced to 9 rules of inference and 10 replacement rules. We will use these rules for the construction of formal proofs. While memorizing these rules and their names will make proofs much easier, the most important thing right now is to be able to “see” how each rule makes sense.

Here is a comprehensive list of the entire set of tools, for your convenience:

Rules of Inference	
Modus Ponens (M.P.)	$  \begin{array}{l}  p \supset q \\  p \\  \therefore q  \end{array}  $
Modus Tollens (M.T.)	$  \begin{array}{l}  p \supset q \\  \sim q \\  \therefore \sim p  \end{array}  $
Hypothetical Syllogism (H.S.)	$  \begin{array}{l}  p \supset q \\  q \supset r \\  \therefore p \supset r  \end{array}  $
Disjunctive Syllogism (D.S.)	$  \begin{array}{l}  p \vee q \\  \sim p \\  \therefore q  \end{array}  $
Constructive Dilemma (C.D.)	$  \begin{array}{l}  (p \supset q) \bullet (r \supset s) \\  p \vee r \\  \therefore q \vee s  \end{array}  $
Simplification (Simp.)	$  \begin{array}{l}  p \bullet q \\  \therefore p  \end{array}  $
Conjunction (Conj.)	$  \begin{array}{l}  p \\  q \\  \therefore p \bullet q  \end{array}  $
Absorption (Abs.)	$  \begin{array}{l}  p \supset q \\  \therefore p \supset (p \bullet q)  \end{array}  $
Addition (Add.)	$  \begin{array}{l}  p \\  \therefore p \vee q  \end{array}  $
Replacement Rules	
Double Negation (D.N.)	$\sim\sim p \equiv^t p$
Commutation (Com.)	$  \begin{array}{l}  (p \vee q) \equiv^t (q \vee p) \\  (p \bullet q) \equiv^t (q \bullet p)  \end{array}  $
Association (Assoc.)	$  \begin{array}{l}  [p \vee (q \vee r)] \equiv^t [(p \vee q) \vee r] \\  [p \bullet (q \bullet r)] \equiv^t [(p \bullet q) \bullet r]  \end{array}  $
Transposition (Trans.)	$p \supset q \equiv^t (\sim q \supset \sim p)$
Material Implication (Impl.)	$p \supset q \equiv^t \sim p \vee q$
De Morgan's Theorems (De M.)	$  \begin{array}{l}  \sim(p \bullet q) \equiv^t (\sim p \vee \sim q) \\  \sim(p \vee q) \equiv^t (\sim p \bullet \sim q)  \end{array}  $
Distribution (Dist.)	$  \begin{array}{l}  [p \bullet (q \vee r)] \equiv^t [(p \bullet q) \vee (p \bullet r)] \\  [p \vee (q \bullet r)] \equiv^t [(p \vee q) \bullet (p \vee r)]  \end{array}  $

Material Equivalence (Equiv.)	$(p \equiv q) \equiv^t [(p \supset q) \bullet (q \supset p)]$ $(p \equiv q) \equiv^t [(p \bullet q) \vee (\sim p \bullet \sim q)]$
Exportation (Exp.)	$[(p \bullet q) \supset r] \equiv^t [p \supset (q \supset r)]$
Tautology (Taut.)	$p \equiv^t (p \vee p)$ $p \equiv^t (p \bullet p)$

## IV. Constructing Formal Proofs Using Rules of Inference and Replacement

A formal proof is a way to show that a conclusion validly follows from a given set of premises. Formal proofs simultaneously use a step-by-step process to a) derive intermediate premises from those that are given toward the conclusion and b) show the justification for each step. The 9 rules of inference and the 10 replacement rules are used to justify the step-by-step moves and allow us to construct formal deductive proofs.

Let's return to the example from the start of the chapter. Below is an example of a finished formal proof. The following paragraphs will explain how to interpret the way formal proofs are written by using this proof as an example.

### Formal Proof Example 1

- |    |                        |                |
|----|------------------------|----------------|
| 1. | $A \bullet B$          |                |
| 2. | $(C \vee A) \supset D$ | $\therefore D$ |
| 3. | $A$                    | 1, Simp.       |
| 4. | $A \vee C$             | 3, Add.        |
| 5. | $C \vee A$             | 4, Com.        |
| 6. | $D$                    | 2,5, M.P.      |

The above example begins with an argument (shown on lines 1 and 2), then adds lines (3-6) to justify the argument and prove its validity. It does so by showing how each new line (3-6) proceeds, one step at a time, by solely utilizing the valid inference and replacement rules we studied in the previous section. Each new line only uses one application of one rule at a time. But the end result is a chain of reasoned steps that justifies the argument and its conclusion. Now, let's break down how this all works in more detail.

Every proof is a proof of an argument. So, we need to back up and look at what the argument is. In this case, the original argument was:

- P1.  $A \bullet B$   
P2.  $(C \vee A) \supset D$   
 $\therefore D$

“(P1)” and “(P2)” indicate the first and second premises. The “ $\therefore$ ” indicates the conclusion. From these we can begin re-writing our proof as:

- |    |                        |                |
|----|------------------------|----------------|
| 1. | $A \bullet B$          |                |
| 2. | $(C \vee A) \supset D$ | $\therefore D$ |

On the far left of the proof, you will see that each line is numbered. Just to the right of the numbers, the next column presents the two premises. It will later include future intermediate premises the proof maker decided to add, and the conclusion, restated.

The far-right column is very, very important. It is the “ledger column.” Right now, we’ve just moved the conclusion to the top of that column. Later that far-right column will contain the information that explains the steps of the argument. Most of this chapter will be dedicated to explaining this ledger column. Here, you’ll notice that the column doesn’t begin until the last row of the premises. In this case, row 2.

As the proof is created, we will add new lines of intermediate premises in the left column and show how such lines are justified by citing the relevant rules in the ledger column. Creating formal proofs is an art. The goal is to proceed toward the conclusion in a way that is elegant and does not clutter things up with extra steps. But there are critical rules.

The first rule is that any step we take, i.e. any new line we add, must be justified by one of the inference or replacement rules we’ve already introduced, or some of the other more complex rules we’ll introduce later.

The second rule is that each step and new line can make only one change at a time and only use one rule at a time.

The third rule is that each step and its justification must match the exact same form as the inference or replacement rule is suited to. For instance, one cannot use the rule “addition” plus the statement “A” to yield “A  $\supset$  B”. That’s just not how that rule works. Addition, to continue, only allows one to start with a line like “A” and get something like “A  $\vee$  \_\_\_” where the blank is filled in with another capital letter. This strict adherence to the *form* of each rule and its application is precisely why formal proofs are, when done correctly, exactly what they claim to be; formal proofs of an argument’s validity.

Looking at the start of our proof above, we can see that our added steps need to ultimately arrive at a “D” to prove our conclusion. In line 2, we see that D follows from a disjunction with an C and A (see the “C  $\vee$  A”). However, the only time we see A in the proof so far is when it is in a conjunction with B (see the “A  $\bullet$  B”). And so, the clever proof-maker might begin by separating A from B. Luckily, the inference rule *simplification* lets us do that from line 1. Recall that “simplification” says that for any (p  $\bullet$  q) set, we are allowed to simplify it to the first conjunct. Deriving A from (A  $\bullet$  B) is an instantiation of “simplification” and is therefore valid. And so, the next step in our proof construction is to create a new line (#3) and write a new intermediate premise “A” there. Then, we use the ledger column to note that the other information in line #3 comes from line #1 and an application of the simplification rule, which is abbreviated “Simp.”

---

1.	A $\bullet$ B	
2.	(C $\vee$ A) $\supset$ D	$\therefore$ D
3.	A	1, Simp.

---

We have created our first new line. Line 3 shows our new bit of information, “A” on the left. The ledger column shows which rule we used to justify our new information with the abbreviation “Simp.” The ledger column also shows which line the rule simplification was applied toward to yield the new output “A”. Here, that was line 1.

Each new line must cite the rule to justifies its new statement. Look at the table of inference and replacement rules. Each one is a little argument. Some, like modus ponens, rely upon two premises to get is conclusion. Others, like simplification only use one premise to deduce its conclusion. When use inference or replacement rules in a proof, we cite exactly as many lines as are needed for the deduction to work. Sometimes, as in simplification, we need only cite one line, as we did in our example proof. At other times, we will need to cite two lines, if the inference step requires.

One may only cite lines above the step on is working on. This is because valid steps can only proceed from information that is already established. Were we to mistakenly cite the same line we are one, or a line latter down in the proof, we’d literally be comiting the fallacy of circular reasoning!

Let’s return to our example. So far, so good. We are getting close. We can see that our conclusion, D, is the consequent that follows from (C  $\vee$  A). And so, it would do well to add a C to our mix. Moreover, we can see that the C appears in line 2 as part of a disjunct. Well, when we covered the rule of *addition* earlier in the chapter we learned that for any variable,

p, it is completely valid to add a second variable as a disjunct. And so, we'll repeat our process and add line 4 to show that we've gotten " $A \vee C$ " from line 3 and the "addition" rule:

---

1.	$A \bullet B$	
2.	$(C \vee A) \supset D$	$\therefore D$
3.	A	1, Simp.
4.	$A \vee C$	3, Add.

---

Again, notice how a new line was added, " $A \vee C$ " and how it was shown as legitimate by the ledger's notation that " $A \vee C$ " follows from application of the addition rule, "Add." to line 3.

Our proof is close to complete. One might now think that combine line 4 with line 2 would give us D via *modus ponens*. The problem is, however, that the propositions labeled "A" and "C" are in different orders as they appear in lines 2 and 4. We cannot take any shortcuts constructing formal proofs. And so, before we do *modus ponens*, we need to restate the info in line 4 by creating a new line and applying the replacement rule of "commutation" which justifies switching the order of conjuncts, yielding:

---

1.	$A \bullet B$	
2.	$(C \vee A) \supset D$	$\therefore D$
3.	A	1, Simp.
4.	$A \vee C$	3, Add.
5.	$C \vee A$	4, Com.

---

Here, we have added another line " $C \vee A$ " and shown how it is legitimately derived from an application of the commutation, "Com." rule to line 4.

Now, we can apply the rule *modus ponens* to both lines 2 and 5 to get "D"! *Modus ponens* is the inference rule that uses two premises, " $(p \supset q)$ " and " $p$ " to derive q. And so, when we apply a specific use of the form of *modus ponens* to get our D, in our ledge column we'll have to refer to both lines involved in that step. *Modus ponens* is one of the inferences that requires utilizing two separate premises. Thus, our completed proof will look like:

---

1.	$A \bullet B$	
2.	$(C \vee A) \supset D$	$\therefore D$
3.	A	1, Simp.
4.	$A \vee C$	3, Add.
5.	$C \vee A$	4, Com.
6.	D	2, 5, M.P.

---

Here, we see our new line 6 that has a "D." The ledger column explains how this "D" was legitimately derived from the application of "*modus ponens*" from both lines 2 and 5.

Boom. Done. Line 6 just reads "D" and that is exactly what the conclusion we set out to prove is. Therefore we are done. Because we have shown how the conclusion, "D," followed from lines 1 and 2 via the application of legitimate rules of inference or replacement, we have shown that "D" is a valid conclusion of our initial argument. In other words, we constructed a formal proof.

We've explained what the numbered lines mean, how we can add intermediate premises to help show the steps needed to reach our conclusion, and how we use the ledger column to show the rule that guarantees the validity of each step. Because each step is valid, we can rest assured that the argument:

P1.  $A \bullet B$

$$\text{P2. } (C \vee A) \supset D$$
$$\therefore D$$

is valid.

Equally important for now is that our example-proof construction showed us that we can only take one step at a time and that we need to justify each and every step.

When we go to add a line of an intermediate premise to our proof, there is no rule that says we must follow a certain order regarding which prior premises or lines we refer to – so long as the occur above the line we are currently adding. Complicated proofs might require some jumping around among the lines above our current location in the proof. In the latter sections we will learn about two more skills, such as “conditional proofs” and “indirect proofs,” which involve fencing off sub-sets of premises, but we’ll cross that bridge when we get there in sections V and VI.

What our example demonstration did not show was exactly how we knew which move to make and when. This is because there is no straightforward and mechanical system for knowing which lines to add and what strategy to pursue in constructing a formal proof. It's a bit of an art. While truth tables are entirely mechanical, formal proofs require some ingenuity. There is some room for variation. Consider the following example of two valid variations of a proof for the same argument:

---

Argument

1.  $M \supset \sim C$
2.  $E \bullet M \quad / \therefore E \bullet \sim C$

To start with, a clever logician might look at the conclusion and identify what components need worked with. Here, we see an E and a  $\sim C$ . Scanning the premises, we see that we'd have the E if we got it from line 2 somehow. And we'd have our  $\sim C$  if we somehow got it from line 1. But this means we'd need an M. Line 2 looks like a good potential place to find that, too. And so, suppose our proof proceeds along the lines of the following variation:

[illegible]

Now, notice how in variation 1,  $M$  was isolated from " $E \bullet M$ " using Com. and Simp. in lines 3 and 4, while  $E$  wasn't isolated until line 6 by using Simp from line 3.

On the other hand, here's an equally valid variation:

---

Validity Proof Variation 2

---

- |    |                    |                               |
|----|--------------------|-------------------------------|
| 1. | $M \supset \sim C$ |                               |
| 2. | $E \bullet M$      | $\therefore E \bullet \sim C$ |
| 3. | $E$                | 2, Simp.                      |
| 4. | $M \bullet E$      | 2, Com.                       |
| 5. | $M$                | 4, Simp.                      |
| 6. | $\sim C$           | 1, 5, M.P.                    |
| 7. | $E \bullet \sim C$ | 5, 6, Conj.                   |
- 

Variation 2 decided to isolate the E and M in a slightly different order than as occurred in variation 1. This is just fine. Sometimes the order doesn't matter for certain steps. Other times, the order does matter, though. For example, note how both arguments used commutation to get the conjunction  $E \bullet M$  switched to  $M \bullet E$  before extracting M with simplification. That's because to keep things consistent, simplification only works with the first part. Therefore, the order of those steps mattered. Likewise, of course, in both variations M had to be isolated before it could be used in the proof as a partner with  $M \supset \sim C$  in the application of *modus ponens*. Again, this order mattered. Whether a formal proof follows the rules and properly shows validity is a closed question. How to get there is not.

**EXERCISES**

Please fill in the ledger column for the following completed proofs:

1.

- |    |               |                     |
|----|---------------|---------------------|
| 1. | $F \supset G$ |                     |
| 2. | $\sim G$      | $\therefore \sim F$ |
| 3. | $\sim F$      |                     |

2.

- |    |                           |                            |
|----|---------------------------|----------------------------|
| 1. | $\sim C \vee \sim B$      |                            |
| 2. | $B \supset C$             | $\therefore \sim B \vee C$ |
| 3. | $\sim (C \bullet B)$      |                            |
| 4. | $B \supset (C \bullet B)$ |                            |
| 5. | $\sim B$                  |                            |
| 6. | $\sim B \vee C$           |                            |

3.

- |    |                                       |   |
|----|---------------------------------------|---|
| 1. | $(F \supset D) \bullet (H \supset J)$ |   |
| 2. | $F \vee H$                            |   |
| 3. | $K$                                   | $\therefore [(K \bullet D) \vee (K \bullet J)]$ |
| 4. | $D \vee J$                            |   |
| 5. | $K \bullet (D \vee J)$                |   |
| 6. | $[(K \bullet D) \vee (K \bullet J)]$  |   |

4.

- |    |                      |                     |
|----|----------------------|---------------------|
| 1. | $\sim A \vee B$      |                     |
| 2. | $F \vee G$           |                     |
| 3. | $\sim G \vee \sim B$ |                     |
| 4. | $\sim F$             | $\therefore \sim A$ |
| 5. | $G$                  |                     |
| 6. | $\sim \sim G$        |                     |
| 7. | $\sim B$             |                     |
| 8. | $\sim A$             |                     |

5.

- |    |                             |
|----|-----------------------------|
| 1. | $[(F \bullet G) \supset C]$ |
|----|-----------------------------|



2.  $G \supset F$
3.  $\sim(\sim G \vee K)$        $\therefore C$
4.  $\sim\sim G \bullet \sim K$
5.  $\sim\sim G$
6.  $G$
7.  $F$
8.  $F \bullet G$
9.  $C$

Construct formal proofs for the provided arguments.

1.
  1.  $A \supset B$
  2.  $\sim(B \bullet A)$        $\therefore \sim A$
2.
  1.  $F$
  2.  $(F \supset G) \bullet (H \vee J)$        $\therefore (G \vee J)$
3.
  1.  $A \supset D$
  2.  $D \supset F$
  3.  $\sim F \vee \sim G$
  4.  $G$        $\therefore \sim A$
4.
  1.  $\sim A \supset (F \vee G)$
  2.  $F \supset D$
  3.  $G \supset H$
  4.  $\sim(D \vee H)$        $\therefore A$
5.
  1.  $[A \supset (B \supset C)]$
  2.  $\sim(C \vee D)$        $\therefore \sim(A \bullet B)$

## V. Conditional Proofs

This section will introduce a technique for deducing conditional statements from existing premises. While our replacement and inference rules enable us to do this already, our new “conditional proof” technique and notation skill will make this process much shorter, simpler, and more manageable.

Conditional “if . . . then . . .” statements are very powerful within our deductive system. Nine out of the 19 rules involve conditionals. In everyday life, it is often critical to establish conditional connections between claims, events, or other statements. Our ability to know things like *B would be true if A were true* can be of crucial importance in our daily reasoning.

Recall that the truth of a conditional does not solely depend on the truth of the antecedent. We can assert  $p \supset q$  without necessarily asserting  $p$ . Often in an argument, it is useful to use existing premises to help us arrive at the claim “if  $p$  were true, then  $q$  would be true” without needing to assert  $p$ . And this conditional itself  $p \supset q$  might then even have a further role in an argument. The method of “conditional proofs” helps us utilize this matter for our deductive proofs.

Learning how to use conditional proofs only requires adding a few new bits of presentation and abiding by a few new rules.

Conditional proofs are sub-proofs. They work by allowing us to introduce a hypothetical assumption into an argument, derive an assumption-specific conclusion, and then “officially” export the fact that the assumption yields

such a conclusion back out one level as an intermediate premise in the bigger argument. The new intermediate premise shows up as a conditional with the assumption as the antecedent and the conclusion as the consequent.

Below is an example of what a finished argument that uses a conditional proof can look like. As with the previous section, the remainder of this section will break down and clarify this example as a way to present the general idea.

---

### Conditional Proof Example 1<sup>2</sup>

1.		
2		
.		
3		
.		
4	$(A \vee B) \supset (C \bullet D)$	
.	$(D \vee E) \supset F$	$\therefore A \supset (F \bullet A)$
5	A	$\therefore F$ (A.C.P.)
.	$(A \vee B)$	3, Add.
6	$C \bullet D$	1, 4, M.P.
.	$D \bullet C$	5, Com.
7	D	6, Simp.
.	$D \vee E$	7, Add.
8	F	2, 8, M.P.
.	$A \supset F$	3-9, C.P.
9	$A \supset (F \bullet A)$	10, Abs.
.		
1		
0.		
1		
1.		

---

## The Anatomy of a Conditional Proof

The above example introduces a conditional proof at line 3 that ends at line 9 and is therefore used to introduce a new intermediate premise at line 10. Looking at the above example, you will see a vertical line that begins at line 3 and runs down all the way to line 9. This line is the *scope line*. The scope line marks out all the steps that are inside the sub-proof.

Moving to the right, after the scope line on line 3, we have “A.” Here, A is the assumption that we’ve strategically introduced for the broader purpose of ultimately getting our conclusion,  $A \supset F$ , from our original given premises 1 and 2. You can introduce any properly written propositional statement as an assumption for a conditional proof – it all depends on what you are trying to prove. The important thing is that you note that the introduction is marked as the “assumption of a conditional proof.” In the ledger column, we have “ $\therefore F$  (A.C.P.)”. After we introduced A, used this notation to justify its introduction by noting it is an Assumption for a Conditional Proof (A.C.P.) that is being introduced specifically to help us get to our target sub-conclusion, in this case F.

Once our conditional assumption is introduced, the remaining lines behind the *scope line* act as a mini-sub proof that get to use all the original premises plus assumption A as sub-premises for its purposes. Thus, you can see lines 4 – 9 using the replacement and inference rules we learned about earlier to ultimately derive F from lines 1, 2, and 3. It is important to also notice how line 9 uses line 2 from the original layer of the proof and line 8 from our sub-proof in combination. This is legitimate; the steps within a sub-proof can pull from the main argument and from its own steps.

Having achieved F at line 9, line 10 then “exports” the conditional “ $A \supset F$ ” back into the main argument. You will see there is no *scope line* on line 10. This is because  $A \supset F$  has been proven as a valid contribution to our intermediate premises. If we assume A, then (given our other premises from lines 1 and 2) we have F, and the ledger note, “3-9, C.P.” documents that this contribution was achieved through the Conditional Proof (C.P.) of lines 3-9.

Notice how line 11 follows the same format as normal proofs from our earlier examples, except that it is using line 10 as its base material for the application of the “absorption” inference rule. This shows how the conclusion from a conditional proof, once exported, becomes a part of the larger argument. And, in our argument, line 10 states our desired conclusion, thus completing the proof as a whole.

Our example walked us through the use of a conditional proof to add an intermediate premise to our argument. The conclusion of a conditional proof, i.e., what we get to export back out, will always be in the form of a conditional with the initial assumption as the antecedent and the desired sub-conclusion as the consequent.

The conditional (assumption  $\supset$  sub-conclusion) is the *only* thing that can be exported from the proof. This is a crucial rule. Any other statement that occurs within the confines of the scope line must stay where it is. The scope line is a one-way barrier. Premises can come in from outside, but nothing can go out.

## Using Multiple Conditional Proofs

Arguments can have multiple conditional proofs. They can occur in serial, as in the following example:

### Serial Conditional Proof Example

1.	$A \vee B$	
2.	$\sim A \vee F$	
3.	$\sim B \vee G$	$\therefore F \vee G$
4.	A	$\therefore F$ (A.C.P.)
5.	$\sim \sim A$	4, D.N.
6.	F	2, 5, D.S.
7.	$A \supset F$	4 – 6, C.P.
8.	B	$\therefore G$ (A.C.P.)
9.	$\sim \sim B$	8, D.N.

- |     |                                   |              |
|-----|-----------------------------------|--------------|
| 10. | G.                                | 3, 9, D.S.   |
| 11. | B $\supset$ G                     | 8 – 10, C.P. |
| 12. | (A $\supset$ F) • (B $\supset$ G) | 7, 11, Add.  |
| 13. | F $\vee$ G                        | 1, 12, C.D.  |

In this example, we have two different conditional proofs, one at 4-6 and the other at 8-9. Each one, “sub-proof 1” and “sub-proof 2,” can draw inferences from lines in “main proof” that have been established above it. However, “sub-proof 2” cannot use any statements from within the scope line of sub-proof 1.

Conditional proofs can also be nested, as shown in example 3, below:

#### Nested Conditional Proof Example

- |     |                                      |  |
|-----|--------------------------------------|--|
| 1.  | A $\supset$ $\sim(\sim D \bullet C)$ |  |
| 2.  | C                                    | $\therefore \sim A \vee D$               |
| 3.  | A                                    | $\therefore D$ (A.C.P.)                  |
| 4.  | $\sim D$                             | $\therefore (\sim D \bullet C)$ (A.C.P.) |
| 5.  | $\sim D \bullet C$                   | 2, 4, Conj.                              |
| 6.  | $\sim D \supset (\sim D \bullet C)$  | 3-5, C.P.                                |
| 7.  | $\sim(\sim D \bullet C)$             | 1, 3, M.P.                               |
| 8.  | D                                    | 6, 7, M.T.                               |
| 9.  | A $\supset$ D                        | 3-8, C.P.                                |
| 10. | $\sim A \vee D$                      | 9, Impl.                                 |

Here, we have a sub-proof (sub-proof 1) at lines 3-8. Within this, we have a “sub-sub” at 4-5. Again, each scope line is a one-way barrier. And so, “sub-sub” can draw from the main proof or the sub-proof 1 lines, but sub-proof 1 is only allowed to take in the official export from the sub-sub. Moreover, sub-sub can only export its conditional one layer back out – in this case, into sub-proof 1. Only the official conditional export of sub-proof 1 gets to appear as an intermediate premise in the main proof, here at line 9.

### *Conditional Proofs for Tautologies*

In addition to supplementing arguments, conditional proofs can be used to prove claims that need no premises at all. Logical tautologies (not to be confused with the replacement rule of the same name) are statements that are always true, no matter what. Thus, they need no premises.

Here is an example of a conditional proof of a tautological statement:

---

### Conditional Proof for Tautology Example

1.		
2	A	$\therefore (A \vee \sim A)$ (A.C.P.)
.	A $\vee$ $\sim$ A	1, Add.
3	A $\supset$ (A $\vee$ $\sim$ A)	1-2, C.P.
.		

---

Notice here how the ultimate conclusion,  $A \supset (A \vee \sim A)$ , does not appear behind the first “ $\therefore$ ” sign as it does with most proofs. Instead, the proof uses that space for stating the conditional proof's conclusion. That's just how it is. Not all tautological statements are as recognizably valid as is  $A \supset (A \vee \sim A)$ . However, being able to prove their truth, lacking any premise whatsoever, can be a useful trick for a logician to be familiar with.

#### EXERCISES

Provide Conditional Proofs for the following arguments:

1.
  1. A
  2.  $B \vee C$       $\therefore \sim C \supset (B \bullet A)$
2.
  1.  $F \vee G$
  2.  $F \supset H$
  3.  $\sim K \vee \sim H$
  4.  $G \supset K$       $\therefore K \equiv G$

## VI. Indirect Proofs

Our second new skill is the use of “Indirect Proofs.” As with conditional proofs, indirect proofs utilize introducing an assumption, scope lines, and sub-arguments to further our main argument. However, in this case what we will be doing is inserting a possible assumption, [sub]proving that the assumption leads to a contradiction, and thereby adding the assumption's negation to the main list of intermediate premises.

Here is an example of a proof that uses an indirect proof. This section will dissect the example to introduce the concept:

---

### Indirect Proof Example 1<sup>3</sup>

1.		
2		
.		
3		
.		
4		
.		
5		
.		
6	$(H \supset I) \bullet (J \supset K)$	
	$(I \vee K) \supset L$	
7	$\sim L$	$\therefore \sim (H \vee J)$
.	$H \vee J$	$\therefore$ (A.I.P)
8	$I \vee K$	1, 4, C.D.
.	$L$	2, 5, M.P.
.	$L \bullet \sim L$	6, 3, Conj.
	$\sim (H \vee J)$	4-7, I.P.

---

The above use of an indirect proof has quite a few commonalities with a conditional proof, so let's flag the differences. The indirect proof is introduced in line 4. Notice the assumption " $H \vee J$ " behind the scope line. Notice, too, the ledger entry. Here, we just have " $\therefore$  (A.I.P.)" to note that line 4 was introduced as an Assumption for an Indirect Proof. Lines 5 and 6 are derivations from lines 1 and 4 then 2 and 5, as is noted in the ledger column.

The final line of the Indirect Proof's sub-proof is important. Here, they've validly derived " $L \bullet \sim L$ " from lines 6 and 3. Well,  $L \bullet \sim L$  is a contradiction and, therefore cannot be true. Because introducing assumption,  $H \vee J$ , is what allowed us to validly derive a contradiction, we are justified in rejecting  $H \vee J$ . Thus, our indirect proof can export  $\sim(H \vee J)$  as its upshot into line 8. Notice both that the scope line has ended, thereby allowing  $\sim(H \vee J)$  to be used in the main argument, and that the ledger column notes the lines of the indirect proof by listing the first through last lines and an "I.P." abbreviation for indirect proof.

All the same principles regarding scope, nesting, and exportation that held for conditional proofs also apply to indirect proofs. Again, the scope line is a one-way barrier, and only one line can be exported out; with indirect proofs, the only legitimate exportation is the negation of the introduced assumption.

3. taken from Copi et al, 2020 p 439

## Proving Tautologies with Indirect Proofs

Just as conditional proofs can be used without premises to prove tautologies, we can also use indirect proofs to provide formal demonstrations of tautological statements. In this case, we introduce the rejection of the tautology as a starting assumption and show that such a rejection leads to contradiction, thus proving the tautology itself.

Below is an example of an indirect proof being used to prove a tautology, in this case,  $G \vee \sim G$ ,

---

Indirect Proof Example 2<sup>4</sup>

1.	$\sim(G \vee \sim G)$	$\therefore$ A.I.P.
2.	$\sim G \bullet \sim \sim G$	1, De M.
3.	$\sim G \bullet G$	2, D.N.
4.	$G \bullet \sim G$	3, Com.
5.	$\sim \sim (G \vee \sim G)$	1-4, I. P.
6.	$G \vee \sim G$	5, D.N.

---

In the above example, the negation of our desired conclusion is immediately introduced to begin an indirect proof leading to its contradiction.

Indirect proofs are a popular way to make a case and/or eliminate untenable propositions. The overall argumentative strategy of *reductio ad absurdum*, which aims to advance a claim by showing that the claim's negation leads to a contradiction is analogous to the use of indirect proof.

A famous example from the history of philosophy is Gottfried Leibniz's (1646-1716) argument for the claim that despite all appearances of evil, we live in the "best of all possible worlds." Leibniz takes as premises the ideas that the world was created by God, and that this God is all-powerful, all-knowing, and all-benevolent. Leibniz rejected the view that our world was somehow flawed or less than the best possible worlds because he thought such a position ran itself into a contradiction when met with his premises about God's perfection. After all, any factor that would force or inspire God to create a lesser world would contradict Leibniz's ideas that God is all-powerful, all-knowing, and all-benevolent – since presumably nothing could force God's hand or diminish God's kindness. Therefore, using a version of indirect proof, Leibniz takes as established the claim that we live the best of all possible worlds.

### EXERCISES

4. from Copi et al, 2020, p 441

Complete indirect proofs for the following arguments.

1.

$$1. S \supset G$$

$$2. S \quad / \therefore G$$

2.

$$1. (A \vee B)$$

$$2. (B \supset D) \quad / \therefore \sim A \supset D$$

3.

$$1. (F \bullet G) \bullet B$$

$$2. \sim H \supset [\sim(F \vee G) \vee \sim B] \quad / \therefore H$$



## 6.

### *Deductive Logic IV: Quantification and Predicate Logic*

SEAN GOULD

## Deductive Logic IV: Quantification and Predicate Logic<sup>1</sup>

### I. Affirmative Singular Predicates

The formal proof methods covered in Chapter V help us construct valid arguments regarding a combination of simple and compound statements. However, if you recall the syllogisms covered in Chapter 3 on Aristotelian syllogisms, then you may see that our formal proofs have a large gap. Chapter V's system doesn't help us express some basic valid arguments.

Consider "Syllogism Example #1":

(P1): All humans are mortal.

(P2): Socrates is a human.

∴ Therefore, Socrates is mortal.

Well, the argument is clearly valid. A Venn diagram would show us as much. But our current method of formal proofs will not. Why? Because here we have three distinct propositions. If we were to try and write them out for a formal proof, we'd be stuck with:

1. A

2. H / ∴ M

The tools we've encountered so far simply do not give us a way grab onto the salient parts that make the argument valid; namely, the features that let categorical propositions (universal affirmative, universal negative, particular affirmative, and particular negative) express what they do about subjects and predicates (recall that the *subject* refers to what a statement is about, and *predicate* refers to whatever descriptive attribution is pinned on to the subject).

### *Affirmative Singular Propositions*

We want to translate Aristotelian Syllogisms into SL so that we can apply our formal proof techniques to a wider pool of arguments. For this, we need *Predicate Logic*. Predicate logic is logic that includes a formal system for talking about subject/predicate relationships. Our first step in gaining this skill will be to learn a system for writing out phrases like "Socrates is a human." Here, "Socrates" refers to an individual subject, while "is human" stands for the predicate term, i.e., the term that gives more illustrative information about the subject. We call statements like this a singular proposition, i.e., a proposition that describes a singular subject. The next sections will show how to construct singular propositions in a notation system with our SL language.

1. This chapter follows Copi, I.M., Cohen, C. & Rodych, V. (2020) Introduction to Logic. 15th Ed. Special Indian Edition. Routledge. Chapter 10, pp 545 - 595. for its organization and general presentation of content. Where direct excerpts are utilized specific citation is given through further footnotes.

Our new system of Predicate Logic will use lower-case letters to refer to the subjects of propositions. So, to start, we could use “s” for Socrates. Typically, the letters *a* through *w* are used to stand for whatever subjects are being discussed and are called *individual constants*. Finally, we will abbreviate the predicate term to a single capital letter, called a *predicate symbol*. We will also put the predicate before the subject.

And so for example, we can use “Hs” to stand for the singular proposition, “Socrates is human.”

To recap:

Hs = Socrates is human.

Where “Socrates” = s

and “is human” = H.

“s” is the individual constant standing for the subject

“H” is the predicate symbol standing for the predicate

The combination of the predicate symbol and the individual constant express the singular proposition.

In a manner analogous to general forms, we might want to replace “Socrates” or “s” with some empty placeholder to simply express the predicate “is human” that could be applied to any particular, singular subject. We will use “x” for that placeholder, also known as an *individual variable*. Henceforth, we will be able to express ideas such as:

Hx = x is human

The entire phrase “Hx” is an example of what is called a *propositional function*. It has a predicate symbol, and individual variable, and when that variable is filled in, with, say, an “s” for Socrates or a “g” for Gandalf, we create a singular proposition. Turning a propositional function into a proposition by filling in the individual variable is called *instantiation*. We replace the variable with an *instance* of its being the case. The predicate “is human” is a *simple predicate*. Simple predicates are propositional functions that could be said of various things, in this example such as Socrates or Yoda, wherein some instances of filling the placeholder yield true propositions while other instances yield false propositions.

The predicate “is human” is just one of a million possible illustrative, simple predicates we might find ourselves representing in our new system. One might say Socrates is wise by using the predicate symbol W for wise, yielding Ws. Or, one might say of some undisclosed x that x is wise; Wx.

Just as we used x as an individual variable, we might also ditch our illustrative predicate for a placeholder that could stand for any simple predicate. A convention is to use Greek symbols, such as “ $\Phi$ ”. And so,  $\Phi s$  means “Socrates is \_\_\_\_\_.” Moreover,  $\Phi x$  means “\_\_\_\_\_ is \_\_\_\_\_.” Blank is blank might be an odd thing to say, but it is helpful to pause upon. As we proceed to learning about categorical syllogisms,  $\Phi x$  helps us get ready to deal with predication in the abstract and not worry about what is being predicated to whatever subject.

Let’s return to our example, Syllogism Example #1:

(P1): All humans are mortal.

(P2): Socrates is a human.

∴ Therefore, Socrates is mortal.

At this point, we have the ability to translate P2 into “Hs”. However, Aristotelian syllogisms require using the terms “all,” “some,” “none” or “some are not.” Our example just uses a universal affirmative as its major premise, but ideally we want to handle any sort of syllogism of the type we covered in chapter 3. To do this, we need to add *quantification* into our system. We will do this by introducing what are called universal and existential quantifiers.

## II. Existential and Universal Quantifiers

The four types of categorical propositions are universal affirmative (all), universal negative (non), particular affirmative (some), and particular negative (some are not). To move toward expressing these categorical propositions, we will introduce existential and universal quantifiers.

Take the sentences “Everything is mortal,” and “All things are mortal.” Unlike our singular proposition, “Socrates is mortal” which only talked about a single subject, “all things are mortal” talks about everything; therefore, we call this a

general proposition. Another way to state “all things are mortal” is to say, “given any  $x$ ,  $Mx$ .” We can reduce the “given any  $x$ ” to just “ $(x)$ ” for short. Thus,

$(x)Mx$  = All things are mortal.

Here, by adding  $(x)$  to  $Mx$  we’ve turned the propositional function into a general proposition through what is called “quantification.” Because the  $(x)$  here means we are talking about any given thing, and thus all things, we call the  $(x)$  symbol the “universal quantifier.”

The universal quantifier lets us get to the notion of your first type of categorical proposition, “universal affirmative” (A) type statements. The idea of the “universal affirmative” can be expressed by the general quantification,  $(x)\Phi x$ .

$(x)\Phi x$  = universal affirmative quantifier – all things are  $\Phi$

In order to say, “no thing is mortal” and express a universal negative, we can negate the predicate function in the affirmative version, thus changing  $(x)Mx$  into the negative  $(x)\sim Mx$ . The more abstract idea of expressing any “universal negative” can be achieved by dropping our illustrative predicate for a predicate variable, yielding:

$(x)\sim\Phi x$  = universal negative quantifier – no things are  $\Phi$

Both  $(x)\Phi x$  and  $(x)\sim\Phi x$  talk about any given thing, whatsoever. They are both talking about “All  $x$ ’s.” The only difference is that one says “all  $x$  are  $\Phi$ ” while the other says “all  $x$  are not  $\Phi$ .” And this is like saying “no  $x$  are  $\Phi$ .”

To talk about the particular affirmative and particular negative categorical propositions, we need a way to say “some  $x$ ” rather than just “for any given  $x$ .” To do this we will introduce a special symbol, “ $\exists$ .” When we have “ $\exists x$ ” we have what is called the “existential quantifier.” Putting an existential quantifier before some illustrative predicate, like mortal, gives us a way to say, “There is some  $x$  that is mortal” or “some  $x$ ’s are mortal.” In predicate logic, we have  $(\exists x)Mx$ . Swapping our  $M$  for a placeholder gives us our third categorical type:

$(\exists x)\Phi x$  = particular affirmative quantifier

To say, “some  $x$ ’s are not \_\_\_\_\_” and express a particular negative, we just add a negation before the predicate function:

$(\exists x)\sim\Phi x$  = particular negative quantifier

Here are the 4 types of general quantifiers all set together for your reference:

$(x)\Phi x$  = universal affirmative quantifier

$(x)\sim\Phi x$  = universal negative quantifier

$(\exists x)\Phi x$  = particular affirmative quantifier

$(\exists x)\sim\Phi x$  = particular negative quantifier

### EXERCISES

Translate the following statements into quantified predicate notation. To do this, first state your abbreviations for the predicate terms.

Sample Problem.                      Sample Solution

Nothing makes sense                       $S$  = makes sense

$(x)\sim(Sx)$

1. Everything counts in large amounts.

2. Some things happen for a reason.

3. Some things do not last.

4. Nothing is permanent.

## III. Subject-Predicate Propositions

We need more than just the four types of general propositions. We need to express full general propositions involving specific predicates and specific subjects, like “All humans are mortal” to express syllogisms such as:

(P1): All humans are mortal.

(P2): Socrates is a human.

∴ Therefore, Socrates is mortal.

We can do this by initially reframing our subject phrases into predicates themselves. For example, take H to stand for being human, as a predicate. Thus  $Hx = x$  is a human. We can rephrase “all humans are mortal” into, “for any given x, if x is a human then x is a mortal.” Using the notation we’ve just learned, we get:

$(x)(Hx \supset Mx)$  = all humans are mortal.

Schematically presented, universal affirmative (A) propositions such as “all \_\_\_\_\_ are \_\_\_\_\_” would look like:

$(x)(\Phi x \supset \Psi x)$  = universal affirmative subject-predicate proposition (A)

Notice we’ve added a second predicate variable,  $\Psi$ , to cover the other term. Since  $(x)\Phi x$  just means “all x’s are \_\_[insert predicate here]\_\_”, we can insert something more complex, and more specific, like “ $(\Phi x \supset \Psi x)$ ” into the blank space. Remember the discussion of universal affirmative statements and existential import as discussed in the chapter on Aristotelian Syllogisms. For a long time in history, our study of logic was complicated by assuming that A statements meant that their subject existed. However, following Boole, it was decided that our inferences went much more smoothly were we to drop this assumption. To say,  $(x)(\Phi x \supset \Psi x)$  is not to affirm that any  $\Phi$ ’d x’s exist. Rather, our A statement just says that for any given thing, if it were to be a  $\Phi$ ’d thing, then it is  $\Psi$ ’d. For example, look at a specific, instantiated universal affirmative claim about the now extinct ivory-billed woodpecker. Suppose we say, “All ivory-bills are birds,” with “ $Ix$ ” as the predicate function for ivory-bill-edness and “ $Bx$ ” being for birds. Then we’d have  $(x)(Ix \supset Bx)$ . And this is true, even though, sadly, no ivory-billed woodpeckers still exist.

Moving on, to cover a universal negative version, we can begin by recognizing that claims such as no “humans are mortal” can be rephrased as something like “for any given x, if x is a human, then x is not mortal” into our notation. This rephrasing helps disclose what the symbolic notation for the general proposition will look like:

$(x)(\Phi x \supset \sim \Psi x)$  = universal negative subject-predicate proposition (E)

The above shows that if  $\Phi$  holds for x, then  $\Psi$  does not. Applying an E statement to human immortality, we’d have  $(x)(Hx \supset \sim Mx)$ ; if something is human, it’s not mortal, which is equivalent to saying, no humans are mortal. Not a true statement, but at least we’ve gained the ability to write it out in a well-formed manner in our notation.

Once again, it is important to note that universal propositions (A) and (E) statements do not assert that anything exists. Moreover, they are not directly opposite statements. Rather, they are *contraries*. Both cannot be true; it can’t be that if any given thing were, it’d both be and not be. However, both can be false at the same time. If some things were not while some were, then either universal claim would be false. “All fish are blue,” and “No fish are blue” are both false in a pool with one red and one blue fish.

When we want to move away from saying, “All x’s are [or are not],” we can start using our existential quantifier – the  $\exists$  symbol. For particular affirmatives and negations, we want to talk about at least one thing, but not everything, and affirm that such a thing exists. And so, to say, “Some humans are mortal,” we might first rephrase the statement as “There is at least one x such that x is both human *and* mortal.” In our notation, the general proposition, about any random set of predicates, would take the form:

$\exists x(\Phi x \bullet \Psi x)$  = particular affirmative subject-predicate proposition (I)

The particular affirmative uses the  $\exists x$  existential quantifier. The existential quantifier affirms that something exists. The above just says there is something for which both predicates hold.

Finally, if we want to say, some humans are not mortal, we can write out,  $\exists x(Hx \bullet \sim Mx)$ . Dropping human and mortal predicates functions for general placeholders, we’d get the general proposition,

$(\exists x)(\Phi x \bullet \sim \Psi x)$  = particular negative subject-predicate proposition (O)

This statement simply says some x’s are such that they are  $\Phi$  and not  $\Psi$ .

For your convenience, here are all four general propositions again:

$(x)(\Phi x \supset \Psi x)$  = universal affirmative subject-predicate proposition (A)

$(x)(\Phi x \supset \sim \Psi x)$  = universal negative subject-predicate proposition (E)

$(\exists x)(\Phi x \bullet \Psi x)$  = particular affirmative subject-predicate proposition (I)

$(\exists x)(\Phi x \bullet \sim \Psi x)$  = particular negative subject-predicate proposition (O)

These general forms are useful because they let us talk about any categorical proposition, not just ones about humans and mortality. We can substitute illustrative predicates for the placeholders as context demands.

By using quantifiers (either  $(x)$  or  $(\exists x)$ ) and two predicate functions ( $\Phi x$  and  $\Psi x$ ) linked with either a conditional or a conjunction, plus an optional negation sign, we have gained a new way to write out categorical propositions that move us one step closer to being able to construct formal proofs of Aristotelian Syllogisms.

Because we are able to translate categorical propositions and predication into our formal language, we can now rewrite Syllogism Example #1.

(P1): All humans are mortal.  
 (P2): Socrates is a human.  
 $\therefore$  Therefore, Socrates is mortal.

Becomes

(P1):  $(x)(Hx \supset Mx)$   
 (P2):  $Hs$   
 $\therefore Ms$

Our next step toward formal proofs will be to add some more inference rules that justify making inferences between lines. The next section will pick up this task in a manner similar to how our existing inference rules were originally explained in Chapter 5.

### EXERCISES

Translate the following A, E, I, or O into quantified predicate notation. To do this, first state your abbreviations for the predicate terms.

Sample Problem.	Sample Solution
All birds are nice.	B = bird
	N = nice
	$(x)(Bx \supset Nx)$

1. Some fish are green.
2. Mules are stubborn.
3. No elephants forget.
4. Some birds do not fly.

## IV. Instantiation, Generalization, and the Quantifier Negation Rule

All of our general quantifiers are about any random  $x$ . Because they are just using variables, such as the “ $x$ ” they literally are “generally speaking.”  $(x)(Mx \supset Lx)$  is a general statement that says all mortals love logic.  $(\exists x)(Mx \bullet Lx)$  is a general statement saying some human out there loves logic.

We sometimes want to move from general statements to speaking about particular instances, such as Socrates. In this case, as you saw in the previous section, we assign a constant, such as “ $s$ ,” to mean Socrates and fill in the  $x$ ’s. If all mortals love logic, and we know Socrates is a mortal, we should be able to say he’s mortal and loves logic, right? Right. To say, if Socrates is mortal, then he loves logic, we’d then write,  $(Ms \supset Ls)$ . Notice, too, how the “ $(x)$ ” at the beginning disappeared because we are no longer talking about any given “ $x$ .” Moving from general statements with individual variables to a specific statement with an individual constant is called “instantiation.” (Think finding an “instance” of the thing).

Sometimes we want to move in the other direction. Suppose we know Socrates is mortal and loves logic ( $Ms \bullet Ls$ ) and we want to perform a “generalization,” and say there is some general person who is mortal and loves logic. First, we’d replace our individual constant,  $s$ , with a variable,  $x$ . We also want to re-introduce a quantifier,  $\exists x$ , to say, “there exists an  $x$  such that . . .”. Thus, we end up expressing this generalization as,  $(\exists x)(Mx \bullet Lx)$ . Again, this just means something exists that is mortal and loves logic.

To summarize, when we move from a variable to a constant, that is called “instantiation.” When we move from a constant to a variable, that is “generalization.” Generalization and Instantiation are important for manipulating

quantified predicates, and we will be exploring rules for their use in the following sections. But first, it is important to detour for a minute and recognize an important fact about how we can negate our general quantifiers.

Suppose I want to say the opposite of  $(x)(Mx \supset Lx)$ . Well, to say not all mortals love logic, I could just write out  $\sim(x)(Mx \supset Lx)$ . However, because our negation is assigned to the quantifier, but not the propositional functions, we don't have a clean way to instantiate the statement. The negation is attached to the quantifier  $(x)$ , and if we drop the quantifier we risk losing the negation and misrepresenting our statement. To get around this issue, we need to think of a different way to negate our quantifier. We need to ensure the negation sign only applies to simple predicates, such as  $Mx$  or  $Lx$ . We call quantified statements that are expressed this way as being *normal-form formulas*. And so, we need to look for ways to express negations in normal form. Well, the negation of "Everything is  $\Phi$ " is "There is something that is not  $\Phi$ ." And so, we can say,

$$\sim(x)\Phi x \equiv^t (\exists x)\sim\Phi x.$$

Similarly, there is not a thing that is not  $\Phi$ , i.e., " $\sim(\exists x)\sim\Phi x$ " is the same as everything is  $\Phi$ . i.e.,

$$\sim(\exists x)\sim\Phi x \equiv^t (x)\Phi x$$

The negation equivalences for "nothing is not  $\Phi$ ," i.e.,  $\sim(x)\sim\Phi$ , and "there is not a thing that is  $\Phi$ " can be equated into normal-form formulas as,

$$\sim(x)\sim\Phi x \equiv^t (\exists x)\Phi x$$

And,

$$\sim(\exists x)\Phi x \equiv^t (x)\sim\Phi x$$

By moving the negation from the quantifier to the propositional function, we will be able to smoothly instantiate and generalize our quantified statements. Therefore, again, it is important to recognize which general quantified statements are negations of each other.

Here, therefore, is a table of quantifications and their negations:

Table 6.IV.1, Quantifier Negations		
$(x)\Phi x$ = universal affirmative proposition	Can have its negation expressed as	$(\exists x)\sim\Phi x$ = particular negative proposition
$(x)\sim\Phi x$ = universal negative quantifier	Can have its negation expressed as	$(\exists x)\Phi x$ = particular affirmative proposition
$(\exists x)\Phi x$ = particular affirmative proposition	Can have its negation expressed as	$(x)\sim\Phi x$ = universal negative proposition
$(\exists x)\sim\Phi x$ = particular negative proposition	Can have its negation expressed as	$(x)\Phi x$ = universal affirmative proposition

Our four traditional subject-predicate categorical propositional forms, the A, E, I, and O statements, can also be negated without needing to add a negation to the quantifier symbol itself. We already saw the contraries of these categorical propositions in the "Square of Opposition." Translating this square into quantified predicate notation, we can yield a table for which categorical propositions can negate each other.

Table 6.IV.2, Categorical Proposition Negations		
$(x)(\Phi x \supset \Psi x)$ = universal affirmative A	Can have its negation expressed as	$(\exists x)(\Phi x \bullet \sim \Psi x)$ = particular negative O
$(x)(\Phi x \supset \sim \Psi x)$ = universal negative E	Can have its negation expressed as	$(\exists x)(\Phi x \bullet \Psi x)$ = particular affirmative I
$(\exists x)(\Phi x \bullet \Psi x)$ = particular affirmative I	Can have its negation expressed as	$(x)(\Phi x \supset \sim \Psi x)$ = universal negative E
$(\exists x)(\Phi x \bullet \sim \Psi x)$ = particular negative O	Can have its negation expressed as	$(x)(\Phi x \supset \Psi x)$ = universal affirmative A

The logical relationship between the quantifier negation pairs and the categorical proposition negations can be established using the rules of inference covered in chapter 5. For example, let's negate the particular negative, O. We know,

$$\sim(\exists x)\Phi x \equiv^t (x)\sim\Phi x$$

And so, filling in the illustrative predicates will compound predicates we have our negated-O and its normal-form formula equivalent,

$$\sim(\exists x)(\Phi x \bullet \sim \Psi x) \equiv^t (x)\sim(\Phi x \bullet \sim \Psi x)$$

Notice how we have consistently replaced the illustrative predicate  $\Phi x$  with the compound  $(\Phi x \bullet \sim \Psi x)$ . Since  $\Phi x$  means  $x$  is \_\_\_\_\_ we are just filling in the blank with  $(\Phi x \bullet \sim \Psi x)$ . Now, take the right side of the biconditional equivalence. If we apply De Morgan's to it, we get,

$$(x)\sim(\Phi x \bullet \sim \Psi x) \equiv^t (x)(\sim\Phi x \vee \sim\sim\Psi x)$$

And we can apply the rule of double negation to the right side to that to get,

$$(x)(\sim\Phi x \vee \sim\sim\Psi x) \equiv^t (x)(\sim\Phi x \vee \Psi x)$$

And we can use Material Implication to say,

$$(x)(\sim\Phi x \vee \Psi x) \equiv^t (x)(\Phi x \supset \Psi x)$$

Hence, condensing this chain of biconditional equivalences, we get,

$$\sim(\exists x)(\Phi x \bullet \sim \Psi x) \equiv^t (x)(\Phi x \supset \Psi x)$$

$(x)(\Phi x \supset \Psi x)$  is thus logically equivalent to the original, non-normal-form formula of  $\sim(\exists x)(\Phi x \bullet \sim \Psi x)$ . Moreover,  $(x)(\Phi x \supset \Psi x)$  is our expression of the universal affirmative A statement. Hence, we have proved that A statements are the negation of O statements.

There might be times when a formal proof involves negating a quantified statement. In this case, we may want to translate it into its normal-form formula for further use. To do so, we can appeal to a new replacement rule:

---

Quantifier Negation Rule Q.N.

One can substitute a negated quantified statement with its equivalent normal-form formula.

Always Valid!

---

Here's a table of negated quantified statements and their equivalent normal-form formulas:

Table 6.IV.3, Quantifier Negation Rule		
negated quantified statement		normal-form formula
$\sim(x)(\Phi x \supset \Psi x)$	$\equiv^T$	$(\exists x)(\Phi x \bullet \sim \Psi x)$
$\sim(x)(\Phi x \supset \sim \Psi x)$	$\equiv^T$	$(\exists x)(\Phi x \bullet \Psi x)$
$\sim(\exists x)(\Phi x \bullet \Psi x)$	$\equiv^T$	$(x)(\Phi x \supset \sim \Psi x)$
$\sim(\exists x)(\Phi x \bullet \sim \Psi x)$	$\equiv^T$	$(x)(\Phi x \supset \Psi x)$

The quantifier negation rule can be used in a formal proof just like any of the other replacement rules so that we may move back and forth between different expressions of the same A, I, E, or O statements.

## V. Formal Proofs of Validity for Aristotelian Syllogisms

Sections II, III, and IV have given us the notational tools for translating Aristotelian Syllogisms into our formal language. And, we just picked up a new quantifier-specific rule, Q.N. Now, we just need to add four more quantifier-specific inference rules and we will be ready to learn about formal proofs of validity for Aristotelian Syllogisms.

To begin, let's re-articulate just what the issue is, and why we have to add these new rules in the first place. To start, normal propositional logic gives us 19 rules for drawing valid inferences and replacements when we are dealing with simple propositions or compounds of propositions [i.e.  $(A \supset B)$  or  $(A \vee B)$  or  $(A \bullet B)$ , etc.]. Our quantifiers give us noncompound statements, too, but these are like "All humans are mortal." Translating this proposition into something like a mere H hides the quantified information inside that gives us the ability to draw inferences. While  $(x)(Hx \supset Mx)$  tells us what we need to know about human mortality, if we take the entire proposition as a whole none of the previously introduced inference rules help us conclude Ms from the two premises  $(x)(Hx \supset Mx)$  and Hs. Why? Although we are close to being able to apply *modus ponens*, Hs does not actually appear in the first premise; only Hx does. Instead, we need to get something like  $Hs \supset Ms$  from the first premise. If we had that, then Ms would follow from Hs.

Generally speaking, for proofs in predicate logic, what we need is a set of rules that helps us move between general quantified statements like  $(x)(\Phi x)$  and  $(\exists x)(\Phi x)$  singular propositions, such as  $\Phi s$ . The following four quantification rules of inference allow just that.

### *Universal Instantiation*

Our first quantification inference rule is called 'Universal Instantiation' and lets us go from a general universal claim to a singular proposition. This rule states, "any substitution instance of a propositional function can validly be inferred from its universal quantification" (Copi et al, 2020: p 469). To explain, let's think about what a universal quantification says: "for any given x,  $\Phi x$ ." Well, if  $\Phi x$  is really true for any given x, then we are justified in giving an instantiation of the x. For example, if any given thing whatsoever is mortal, and Socrates is a potentially given thing, then from  $(x)Mx$  we are welcome to conclude Ms, i.e., that Socrates is mortal. And this sort of inference holds valid for any use of  $(x)\Phi x$  and any individual constant we want to give it. Rather than the individual variable "x" we will use "v" to stand as some individual constant, which could really be "s" or "a" or whatever. Schematically we can say:



---

Universal Instantiation (U.I.)

$$\frac{(x)(\Phi x)}{\therefore \Phi v}$$

Always Valid!

---

With Universal Instantiation, we take the phrase “any given  $x$ ” up on its offer and say, okay, how about “ $v$ ”, let’s give it that.

In the above example and those that follow, we will stick to single function statements in our formal display. Such as, U.I. allows  $\Phi v$  from  $(x)(\Phi x)$ . However, any combination of propositional functions can go after the quantifier. For example, U.I. also allows  $(\Phi v \supset \Psi v)$  from  $(x)(\Phi x \supset \Psi x)$ , or any other replacement provided the propositional functions, and logical operators remain consistent following the application of the rule.

Universal Instantiation just so happens to be all that we need to write out a formal proof of our example Aristotelian syllogism about Socrates’s human mortality. Here now is the full proof as an example of Universal Instantiation in action.

---

Proof of Syllogism Example # 1

1.	$(x)(Hx \supset Mx)$	
2.	$Hs$	$\therefore Ms$
3.	$Hs \supset Ms$	1, U. I.
4.	$Ms$	3, 2, M.P.

---

Here you see the application of Universal Instantiation on line 3. Notice also that we are treating the singular propositions  $Hs$  and  $Ms$  just like we treat other propositions, such as  $A$  or  $B$ , as we proceed through the remaining inferences. In this case, we applied *modus ponens* to achieve our conclusion.

## Universal Generalization

Our next inference rule works like Universal Instantiation but in the other direction. Here, we go from a special sort of particular back to the universal. Suppose we pluck an arbitrary individual out of thin air, call it “ $y$ ,” *and not knowing anything else about it*, find out through other inferences that  $\Phi y$  holds true to it. Because we know nothing about  $y$ , it is valid to acknowledge that  $y$  could just as well have been “any given thing.” Well, if any given thing could be  $\Phi$ , then everything is  $\Phi$ . This brings us to our second quantifier inference rule; universal generalization.

The following example (adapted from Copi, p. 470) will help further illustrate and explain just what Universal Generalization is and how it works. Suppose we want to build a proof for the following argument:

(P1): All humans are mortal.  
(P2): All Aztecs are human.  
 $\therefore$  All Aztecs are mortal.

First, we can write this out with our new notation system, where Hx functions for Human, Mx mortal, and Ax predicates Aztec to its subject. Thus, we have:

---

Syllogism Example 2:	
1.	$(x)(Hx \supset Mx)$
2.	$(x)(Ax \supset Hx) \quad / \therefore (x)(Ax \supset Mx)$

---

Applying Universal Instantiation and our existing inference rules to this argument gets us close to the desired conclusion, but not close enough. Here's as far as we get with Universal Instantiation and hypothetical syllogism:

---

Partial Proof of Syllogism Example 2:	
1.	$(x)(Hx \supset Mx)$
2.	$(x)(Ax \supset Hx) \quad / \therefore (x)(Ax \supset Mx)$
3.	$Hy \supset My \quad 1, \text{U.I.}$
4.	$Ay \supset Hy \quad 2, \text{U.I.}$
5.	$Ay \supset My \quad 3, 4, \text{H.S.}$

---

So far, we've seen that Universal Instantiation lets us plug any particular instance into the universal affirmatives provided in lines 1 and 2. But, suppose for a particular instance we use unknown individual "y." In Syllogism Example #2, the rule Hypothetical syllogism lets us get line 5 from lines 3 and 4. And so, we affirm that if unknown y is Aztec, then y is human. However, our conclusion is not about "unknown individual y." Rather, it is about all Aztec people. Needless to say, because "y" is so vague, it could represent any given person. Nothing has been said about y that means it couldn't just as well be anyone. All we've done was apply the premised claims that for any given thing, if that thing is human then it's mortal, and if that thing is Aztec then it's human. Because any substitution from U.I. is true, then all substitutions must be true. Hence, we can move from y back to  $(x)\Phi x$ .

Our new inference rule, Universal Generalization follows the above line of reasoning. Here is the general schema:

---

Universal Generalization (U.G.)	
	$\Phi y$
	$\therefore (x)(\Phi x)$
Always Valid (if you follow the rules)!	

---

There is an important rule we must follow. Jumping from "y" to the universal quantifier only works when "y" is any given individual. Hence, we will only use "y" when it's been introduced as the assumed "any given random thing." For our purpose, this would be when y is introduced through U.I. in the first place or as an assumption in a conditional proof. I mean, we cannot reasonably infer that just because Socrates is Greek, then everyone is Greek. Our individual constant must represent an "unknown thing."

Syllogism Example #2, shows how we can properly use U.G. to finish a proof.

---

Complete Proof of Syllogism Example 2:

- |    |                      |                                 |
|----|----------------------|---------------------------------|
| 1. | $(x)(Hx \supset Mx)$ |                                 |
| 2. | $(x)(Ax \supset Hx)$ | $\therefore (x)(Ax \supset Mx)$ |
| 3. | $Hy \supset My$      | 1, U.I.                         |
| 4. | $Ay \supset Hy$      | 2, U.I.                         |
| 5. | $Ay \supset My$      | 3,4, H.S.                       |
| 6. | $(x)(Ax \supset Mx)$ | 5, U.G.                         |
- 

## Existential Generalization

The preceding rules cover generalization and instantiation for universal statements. We also need two inference rules that let us move back and forth between singular propositions, e.g.,  $\Phi v$ 's, and quantified functions,  $\exists x$ 's.

Recall that  $\exists x(\Phi x)$  is true if at least one individual can have  $\Phi$  predicated to it. And so, if we have an instance of  $\Phi v$ , (where  $v$  is any individual constant), we've satisfied the condition for  $\exists x(\Phi x)$ . We can articulate this valid inference via the Existential Generalization rule.

---

Existential Generalization (E.G.)

$$\frac{\Phi v}{\therefore (\exists x)(\Phi x)}$$

Always Valid!

---

Existential Generalization is, therefore simply the rule that if we have at least one specific example of something, we can say that there exists an example of that thing.

## Existential Instantiation

It may help to introduce our final quantification rule with a less abstract example. Suppose you have ordered a pizza, which you momentarily leave on the counter. When you return, you find the pizza gone. Being a keen logician, you say, "we know there exists something which stole our pizza. Let's call that "something 'v.'" Maybe not a conclusive step, but so far this is a deductively valid way to proceed.

The rule of Existential Instantiation formalizes the reasoning that occurs in examples like the above. If we know there exists something with a certain predicate, i.e.,  $(\exists x)(\Phi x)$ , then we are warranted to give some arbitrary name to "that thing which exists" in order to talk about it further.

The important caveat here is that the name we give it must not occur anywhere prior in our argument. Return to the pizza example. No matter how guiltily your dog, Shadow, looks at you, without further evidence, you are not entitled to just say, "we know there exists something which stole our pizza. Let's call that something . . . Shadow. Bad dog, Shadow."

The name we use needs to be neutral in our context, lest we accidentally contaminate our argument with any invalid inferences. In other words, the name cannot be used earlier in the argument.

Thus, our formal rule is,

---

Existential Instantiation (E.I.)
$\frac{(\exists x)(\Phi x)}{\Phi v}$
(where v is any individual constant <i>with no previous use in the argument</i> )
Always Valid (if you follow the rules!)

---

With this rule in place, let's review our pizza problem to see exactly how much trouble we get into if we break the, "only use new constants" rule.

Let's use, Px to predicate "ate the pizza" to something and Sx to predicate "being Shadow-dog" to something. And so, our givens at the start of the argument are that something ate the pizza, and something [something else potentially] is Shadow:

1.  $(\exists x)Px$
2.  $(\exists x)Sx$        $\therefore (\exists x)(Px \bullet Sx)$ . [Invalid!]

Now, our invalid argument tries to reach its unjust conclusion by using the same constant twice in repeated applications of Existential Instantiation. The unjustness occurs when we assume or let the idea slip into our argument that the same thing which is Shadow is what ate the Pizza. Here's the formal argument with the ledger column and an added "notes" column to help clarify the steps and missteps:

---

Invalid Example #1		
1. $(\exists x)Px$		
2. $(\exists x)Sx$	$\therefore (\exists x)(Px \bullet Sx)$	Invalid Conclusion!
3. Pa	1, E.I.	Okay step
4. Sa	2, E.I.	NO! "a" is already in use!
5. $Pa \bullet Sa$	3,4, Conj.	Okay step, but tainted by error on line 4
6. $(\exists x)(Px \bullet Sx)$	5, E.G. v	Invalid, due to error on line 4

---

Because the second use of E.I. on line 4 uses a constant that previously occurs, rather than a new one, the argument

then invalidly proceeds to identify "a" with the pizza eater. Had the argument followed the rules, it couldn't have reached the conclusion.

Here is an example of a valid syllogistic proof using two particular statements and a correct application of the Existential Instantiation rule:

Syllogism Example #3:

- P1: No mammals are cold-blooded  
P2: Some egg-layers are mammals

∴ Some egg-layers are not cold-blooded

Let M, C, and E predicate mammal, cold-blooded, and egg-laying. Here is a formal proof of our argument in predicate logical notation (with explanatory notes):

---

### Syllogism Example #3

1.	$(x)(Mx \supset \sim Cx)$		
2.	$(x\exists)(Ex \bullet Mx)$	/ ∴ $(x\exists)(Ex \bullet \sim Cx)$	
3.	$Ea \bullet Ma$	2, E.I.	We get to randomly assign “a”
4.	$Ma \bullet Ea$	3, Com.	
5.	$Ma$	4, Simp.	We can re-use “a” when using U.I.
6.	$Ma \supset \sim Ca$	1, U.I.	
7.	$\sim Ca$	5,6, M.P.	.
8.	$Ea$	3, Simp	.
9.	$Ea \bullet \sim Ca$	7,8, Conj.	
10.	$(\exists x)(Ex \bullet \sim Cx)$	9, E.G.	

---

The above syllogism is valid, and thanks to the Platypus, it's sound as well.<sup>2</sup>

## VI. Asyllogistic Inferences

The method of assigning predicates to things and using our logical operators of  $\supset$ ,  $\vee$ ,  $\bullet$ ,  $\equiv$ , and  $\sim$  can help us formalize arguments beyond Aristotelian syllogisms. Inferences that utilize predicates but move beyond Aristotelian syllogism are *asyllogistic inferences*. With our new system, we can talk about how different predicated objects related to each other, beyond just how objects might or might not fit into categories.

Here is an example of using predicate logic beyond Aristotelian syllogisms. I used to work in an outdoor shop that repaired gear. Behind the cash register, there was a sign that said:

A job can be fast, cheap, or good, but not all three.  
 If a job is fast and good, it won't be cheap  
 If a job is good and cheap, it won't be fast.  
 If a job cheap and fast, it won't be good.

We can formalize the above axioms into asyllogistic predicates and make asyllogistic inferences. For example, if we use  $Fx$ ,  $Cx$ ,  $Gx$ , and  $Jx$  for Fast(x), Cheap(x), Good(x), and Job(x), we can rewrite our example as follows:

- The form of Syllogism Example #3, is called “Ferio” in its medieval nomenclature. It has EIO statements in that order, and the first, i.e., the major premise, has the conclusion's predicate as its predicate and the middle term (the one shared by the premises but not the conclusion) as its subject, while the second, i.e., minor premise, has the middle term as its predicate.

---

### Translation of “the Sign”

$(x)\{Jx \supset [(Fx \vee Cx) \vee Gx] \bullet \sim[(Fx \bullet Cx) \bullet Gx]\}$   
 $(x)\{Jx \supset [(Fx \bullet Gx) \supset \sim C]\}$   
 $(x)\{Jx \supset [(Gx \bullet Cx) \supset \sim Fx]\}$   
 $(x)\{Jx \supset [(Cx \bullet Fx) \supset \sim Gx]\}$

A job can be F, C, or G, but not all three.  
 If a job is fast and good, it won't be cheap.  
 If a job is good and cheap, it won't be fast.  
 If a job is cheap and fast, it won't be good.

---

Had a customer come in and asked for a good job, I could use the signs axioms and predicate logic to formally demonstrate the validity of telling them that a good job would either be a) not cheap or b) not fast!

---

### Asyllogistic Formal Proof #1

1.	$(x)\{Jx \supset [(Fx \vee Cx) \vee Gx] \bullet \sim[(Fx \bullet Cx) \bullet Gx]\}$	
2.	$(x)\{Jx \supset [(Fx \bullet Gx) \supset \sim C]\}$	
3.	$(x)\{Jx \supset [(Gx \bullet Cx) \supset \sim Fx]\}$	
4.	$(x)\{Jx \supset [(Cx \bullet Fx) \supset \sim Gx]\}$	$\therefore (x)[(Jx \bullet Gx) \supset (\sim Cx \vee \sim Fx)]$
5.	$Jy \bullet Gy$	$\therefore (\sim Cy \vee \sim Fy)$ (A.C.P.)
6.	$Jy$	5, Simp
7.	$Gy \bullet Jy$	5, Com.
8.	$Gy$	7, Simp.
9.	$Jy \supset [(Cy \bullet Fy) \supset \sim Gy]$	4, U.I.
10.	$(Cy \bullet Fy) \supset \sim Gy$	6, 9, M.P.
11.	$\sim \sim Gy$	8, D.N.
12.	$\sim(Cy \bullet Fy)$	10, 11, M.P.
13.	$\sim Cy \vee \sim Fy$	12, De.M.
14.	$(Jy \bullet Gy) \supset (\sim Cy \vee \sim Fy)$	5 – 13, C.P.
15.	$(x)[(Jx \bullet Gx) \supset (\sim Cx \vee \sim Fx)]$	14, U.G.

---

In the above proof, you will notice a conditional proof begins on line 5, where we suppose “y” is both a job and is good. Conditional proofs are totally fine to use within predicate logic, as long as we mind the rules. From this particular conditional proof and our prior premises, we arrive at the conclusion that it will either be not cheap or not fast. At line 14, we then export the conditional as a validly derived claim.

The next step is important to pay attention to, as it involves the complicated U.G. Since our instantiated job “y” was any given job and introduced into the main proof at line 14 from a mere assumption – with no other information assumed about “y” (or predicated onto “y”) we can apply Universal Generalization – after the conditional proof is done – to conclude that for anything, if it's a job that is good, then it either won't be cheap or won't be fast.

We can only use U.G. for constants introduced in a conditional proof after the statement has been exported because using U.G. beforehand would result in creating an invalid universal claim about just a part of the overall conditional. For example, applying U.G. to line 8 would have given us the nonsense claim  $(x)(Gx)$ , i.e., all things are good from the mere assumption that “y” was a good job.

Predicate logic can also be used with indirect proofs, as shown in the example, below:

Indirect Proof with Predicate Logic Example		
1. $(x)(Mx \supset \sim Cx)$		
2. $(\exists x)(Ex \bullet Mx)$	$\therefore (x)(Ex \bullet \sim Cx)$	
3. $  (x)(Ex \supset Cx)$	$\therefore$ (A.I.P.)	Here we start an indirect proof
4. $  Ey \bullet My$	2, E.I.	New use of "y" following E.I. rules
5. $  My \supset \sim Cy$	1, U.I.	Any symbol, include y can be used with U.I.
6. $  My \bullet Ey$	4, Com.	
7. $  My$	6, Simp.	
8. $  \sim Cy$	5, 7, M.P.	
9. $  Ey$	4, Simp.	
10. $  Ey \supset Cy$	3, U.I.	Any symbol can be used with U.I.
11. $  Cy$	9, 10, M.P.	
12. $  Cy \bullet \sim Cy$	8, 11, Conj.	Yes, it's a contradiction, but that's the point.
13. $\sim (x)(Ex \supset Cx)$	3-12, I.P.	Export the negated I.P. assumption
14. $(\exists x)(Ex \bullet \sim Cx)$	13, Q.N.	Apply Q.N. to return to the normal-form formula.

Asyllogistic inferences use all the rules of inference that formal proofs utilize. The additional skill that is required in their application to arguments lies beyond formal logic. That skill is of accurate interpretation of propositions into formal notation so that the intended meaning of the propositions is preserved.

For example, consider the following example – “oysters and clams are delicious” (from Copi, 2020: 482). If we take this to mean oysters are delicious and clams are delicious, we need to be careful in what our symbols represent. We can take it as a conjunct of propositions, one about oysters, and one about clams,

$$(x)[(Ox \supset Dx)] \bullet (x)[(Cx \supset Dx)]$$

Or, we can take it as non-compound claim that if a thing is either an oyster or a clam, then it'd be delicious,

$$(x)[(Ox \vee Cx) \supset Dx]$$

But we don't want to just say,

$$(x)[(Ox \bullet Cx) \supset Dx]$$

Because this statement entails that anything that is delicious is somehow both an oyster and clam at the same time. Seems fishy to me. Moving effectively from thought to language is an art regardless of whether or not the language is English, Spanish, Latin, Swahili, Hawaiian, Arabic, Mandarin, or a formal logical notation system.

### EXERCISES

Complete the ledger column for the following arguments.

1.

1.  $(x)(Ax \supset Bx)$
2.  $(x)(Bx \supset Cx)$      $\therefore (x)(Ax \supset Cx)$
3.  $Aa \supset Ba$
4.  $Ba \supset Ca$
5.  $Aa \supset Ca$
6.  $(x)(Ax \supset Cx)$

2.

1.  $(x)(Fx \supset Gx)$
2.  $(x)(Gx \supset \sim Hx)$      $\therefore \sim (\exists x)(Fx \bullet Hx)$
3.  $Fa \supset Ga$
4.  $Ga \supset \sim Ha$
5.  $Fa \supset \sim Ha$
6.  $(x)(Fx \supset \sim Hx)$
7.  $\sim (\exists x)(Fx \bullet Hx)$

3.

1.  $(\exists x)(Sx \bullet Gx)$
2.  $(x)(Gx \supset Mx)$      $\therefore (\exists x)(Sx \bullet Mx)$
3.  $Sa \bullet Ga$

4.  $Ga \bullet Sa$
5.  $Ga$
6.  $Ga \supset Ma$
7.  $Ma$
8.  $Sa$
9.  $Sa \bullet Ma$
10.  $(\exists x)(Sx \bullet Mx)$

4.

1.  $(x)(Rx \vee Dx)$
2.  $(\exists x)(\sim Rx) \quad / \therefore (x)(Dx)$
3.  $\sim Ra$
4.  $Ra \vee Da$
5.  $Da$
6.  $(\exists x)(Dx)$

5.

1.  $(x)(Hx \supset Px) \bullet (x)(Sx \supset Wx) \quad / \therefore (x)[(Hx \vee Sx) \supset (Px \vee Wx)]$
2.  $(Ha \supset Pa) \bullet (Sa \supset Wa)$
3.  $| Ha \vee Sa$
4.  $| Pa \vee Wa$
5.  $(Ha \vee Sa) \supset (Pa \vee Wa)$
6.  $(x)[(Hx \vee Sx) \supset (Px \vee Wx)]$

Complete formal proofs for the following arguments.

1.

1.  $(x)(Mx \supset \sim Px)$
2.  $(\exists x)(Fx \bullet Px) \quad / \therefore (Ex)(Fx \supset \sim Mx)$

2.

1.  $(\exists x)[Gx \supset (Hx \vee Jx)]$
2.  $(x)(Hx \supset \sim Jx) \quad / \therefore (x)[Gx \supset \sim(Hx \bullet Jx)]$

Translate the following arguments into predicate logic and then construct formal proofs of their validity.

1.

- (P1) All skunks are mammals.  
 (P2) No invertebrates are mammals.  
 No invertebrates are skunks.

2.

- (P1) Any given teapot is neither here nor there.  
 (P2) Some folks are there.  
 Some folks are not teapots.

3.

- Some hobbits are Bullroarers, and all Bullroarers are tall. And so, some hobbits are tall.

## VII. Proving Invalidity with Predicate Logic

It only takes one bad oyster to disprove the claim that all oysters and clams are tasty. We have seen that proving validity using quantifiers and predicate logic works along the lines of the formal proofs we explored above and in Chapter 5. It is also possible to assign truth values to the premises and the conclusion and use the Truth Table technique developed in Chapter 4.



One of the strengths of deductive logic is that we can use its stripped-down ‘formal-relationship-only’ approach to investigate the validity and invalidity of any given argument’s logical analogy. If two arguments have the exact same logical structure and we can prove one to be invalid, then since invalidity depends on structure alone, we’d know the other is also invalid. This can be very useful when the original argument in question is clouded by vested interest, bias, complications, or whatnot.

Refutation by logical analogy can be achieved in a special way when dealing with syllogisms and quantified propositions; the use of the counter-example. Copi et al put the idea very well,

An argument involving quantifiers is valid if, *and only if*, it is valid no matter how many individuals there are, provided there is at least one. So, an argument involving quantifiers is proved invalid if there is a possible universe or *model* containing at least one individual such that the argument’s premises are true and its conclusion false *of that model* (2020, p 477).

In other words, just as with Venn diagrams, if we can find conceptual space for an individual counterexample to an argument, we debunk the entire deductive argument. Consider the invalid syllogism from earlier in the text:

No M are P.  
Some S are not M.  
∴ Some S are P.

A Venn diagram would show where a counter-example can be found. Constructing a Venn Diagram shows that it is possible to preclude anything from being in the M/P overlap zone (per premise 1), have an object in the S circle that is outside the M circle (per premise 2), and yet still not be forced to fill in an object in the S/P overlap (per conclusion). Again, for an argument to be valid, the premises must *necessitate* the conclusion – force it on us. With true premises and a false conclusion, this argument is invalid.

In our new quantified notation, we can see the same thing:

(P1.)  $(x)(Mx \supset \sim Px)$  – Universal Negative (E): no M are P  
(P2.)  $(\exists x)(Sx \bullet \sim Mx)$  – Particular Negative (O): Some S are not M  
∴  $(\exists x)(Sx \bullet Px)$  – Particular Affirmative (I): Some S are P

General propositions, like universal affirmative  $(x)\Phi x$  or  $(x)\sim\Phi x$  universal negatives, are only true if they hold of all the individuals a world with any possible number of individuals. This means they are true in scenarios where there is only one thing, true in scenarios with 2 things, 3 things, etc.

And so, in a world with just one thing, called “a”

$(x)(\Phi x) \equiv^t \Phi a \equiv^t (\exists x)(\Phi x)$

The first part states the universal quantifier. The second gives the predicate to the only thing that might exist. The third part says the  $\Phi$ ’d thing does exist.

In a world with two things, a and b, we know the following two principles must be true:

$(x)(\Phi x) \equiv^t [\Phi a \bullet \Phi b]$   
and  $(\exists x)(\Phi x) \equiv^t [\Phi a \vee \Phi b]$

The first part, here, deals with universal statements and says everything, including a and b are  $\Phi$ . The second deals with existential statements and says  $(\exists x)(\Phi x)$  is logically equivalent to saying that at least one of the two things that exist is a  $\Phi$ ’d thing, but we don’t know which one.

In a world with three things, we have

$(x)(\Phi x) \equiv^t [\Phi a \bullet \Phi b \bullet \Phi c]$   
and  $(\exists x)(\Phi x) \equiv^t [\Phi a \vee \Phi b \vee \Phi c]$

This says the same as above but with a third object in our possible world. If everything is  $\Phi$ , then everything listed out is  $\Phi$ . Dealing with the next line, if we know at least one thing is  $\Phi$ , then we know at least one thing out of the entire list of things is  $\Phi$ , but we don’t know which one.

To cut the matter short and write out our rule for any given number of things, we’ll use “. . . n” to represent the continuation from c to some arbitrary amount “n”. Hence:

$(x)(\Phi x) \equiv^t [\Phi a \bullet \Phi b \bullet \Phi c \bullet \dots \bullet \Phi n]$   
and  $(\exists x)(\Phi x) \equiv^t [\Phi a \vee \Phi b \vee \Phi c \vee \dots \vee \Phi n]$

When we want to prove invalidity of a bad syllogistic inference, we can start by assuming a world where just one thing exists and see if we can assign true values to the premises but a false one to the conclusion. If we cannot, we can try to use a two-world model and see if we can have true premises and a false conclusion. The upshot is that when we are dealing with quantified statements, we logically allowed to look for things that would serve as counterexamples to our conclusion.

We find counter-examples by trying to assign true values to instances of the premises and a false value to the conclusion here (while being consistent in our truth-value attributions, of course), thereby showing that the argument is not valid. Rather than by using a Venn Diagram, we can assume a world where only one thing exists and assign a constant to the “x” in the argument in both the premises and the conclusion. If the argument holds true for that, we can see if there is space for a second individual which falls outside the conclusion, or a third, etc.

In the following argument, we are told that one individual exists, but we can invalidate the conclusion by appealing to a second individual that stands as a counterexample.

---

Invalid Example #2  
P1.  $(x)(Mx \supset \sim Px)$   
P2.  $(\exists x)(Sx \bullet \sim Mx)$   
 $\therefore (x)(Sx \supset Px)$

---

Now, can we make the premises true but the conclusion false? Yes, we can! Just because we are told that there is an individual that is S and  $\sim M$  doesn't mean that there might also be and second individual thing that is S and M. And we know that if a thing is M then it is  $\sim P$ . So, our S thing would be M and thus  $\sim P$ . But this would give us  $(\exists x)(Sx \bullet \sim Px)$ , which is the same as  $\sim(x)(Sx \supset Px)$ . And so, we have a situation where the premises are true, but the conclusion is false. Hence, the argument does not force the conclusion and is therefore invalid.

Sometimes it's not easy to pull a counter-example from thin air. A much more mechanical way to proceed is to systematically search for options where the conclusion is false, but the premises are true by using a truth table. We initially try to demonstrate this by utilizing a truth table where our premises are true, but the conclusion is false in an imagined world where only one individual exists. For this example, we will walk through building up the complete truth table one step at a time. First, let's lay out the premises and conclusions as if M, P, and S only described “a”.

Truth-Table Proof of Invalidity for Invalid Example #2					
Ma	Pa	Sa	$Ma \supset \sim Pa$	$Sa \bullet \sim Ma$	$Sa \supset Pa$

Next, let's assign a false value to our conditional-form conclusion statement:

Ma	Pa	Sa	$Ma \supset \sim Pa$	$Sa \bullet \sim Ma$	$Sa \supset Pa$
					F

From here, we use the truth conditions for the conditional to force the values for Sa and Pa. Since a conditional is only false if the antecedent is T while the consequent is false, we have:

Ma	Pa	Sa	Ma	$\supset$	$\sim$	Pa	Sa	$\bullet$	$\sim$	Ma	Sa	$\supset$	Pa
	F	T									T	F	F

Now, we use those assigned values to backfill the other instance of Sa and Pa where they occur.

Ma	Pa	Sa	Ma	$\supset$	$\sim$	Pa	Sa	$\bullet$	$\sim$	Ma	Sa	$\supset$	Pa
	F	T				F	T				T	F	F

Next, we start filling in the premises in an attempt to make them all true. Let's start with the first premise. If Pa is F, the  $\sim$ Pa is T. Hmmm. A conditional with a true consequent, here  $\sim$ Pa, is true regardless of the antecedent. This leaves Ma undetermined. Let's try the second premise, then. Here, Ma must be false for  $\sim$ Ma to be true, which is the requirement of the conjunct. And so, we have:

Ma	Pa	Sa	Ma	$\supset$	$\sim$	Pa	Sa	$\bullet$	$\sim$	Ma	Sa	$\supset$	Pa
	F	T				T	F	T	T	<b>T</b>	F	T	F

Now we have the value for Ma, so we can fill out the rest of the table:

Truth-Table Proof of Invalidity for Invalid Example #2													
Ma	Pa	Sa	Ma	$\supset$	$\sim$	Pa	Sa	$\bullet$	$\sim$	Ma	Sa	$\supset$	Pa
F	F	T	F	T	T	F	T	T	T	F	T	F	F

And voila! We've given a consistent description of a possible scenario where the premises are true, but the conclusion is false.

This method works for worlds with two individuals as well. But the truth table gets quite big, fast. Recall that for two individuals:

$$(\forall x)(\Phi x) \equiv^t [\Phi a \bullet \Phi b]$$

$$\text{and } (\exists x)(\Phi x) \equiv^t [\Phi a \vee \Phi b]$$

And so, this tells us we need to take whatever statement is being quantified, such as  $(Ma \supset \sim Pa)$ , and adjust it accordingly depending on the quantifier ( $\forall$ ) or ( $\exists$ ). For example compare "Single Item Instantiation" with "Two-Item Instantiation" of our example argument as given in the table below:

---

Invalid Example #2 Instantiated in Single Individual and Two-Individual Worlds

Single Item Instantiation

P1.  $Ma \supset \sim Pa$   
 P2.  $Sa \bullet \sim Pa$   
 $\therefore Sa \supset Pa$

Two-Item Instantiation

P1.  $(Ma \supset \sim Pa) \bullet (Mb \supset \sim Pb)$   
 P2.  $(Sa \bullet \sim Pa) \vee (Sb \bullet \sim Pb)$   
 $\therefore (Sa \supset Pa) \bullet (Sb \supset Pb)$

---

The truth table ends up having many columns (to save space on the page, we've cut out the columns that list the variables). In our new two-item world it is possible that either a or b might end up being the instantiation of our counterexample. In the following fleshed out truth-table, it just so happens that the item that is S and  $\sim P$  is the "b" item:

Truth Table Proof of Invalidity for Invalid Example #2 in a Two-Item World																								
(Ma ⊃ ~ Pa) • (Mb ⊃ ~ Pb)								(Sa • ~ Ma) ∨ (Sb • ~ Mb)								(Sa ⊃ Pa) • (Sb ⊃ Pb)								
T	T	T	F	<b>T</b>	F	T	T	F	T	F	F	T	<b>T</b>	T	T	T	F	T	T	T	<b>F</b>	T	F	F

Luckily, because it only takes one possible, consistent model of a counterexample to prove invalidity, our truth-table only needs one row.

**EXERCISES**

Use the Truth-Table Method to show the following arguments are invalid:

1.

(P1)  $(\exists x)(Fx \bullet Kx)$   
 $\therefore (x)(Fx \supset Kx)$

2.

(P1)  $(x)(Hx \supset Mx)$   
 (P2)  $(\exists x)(Hx \vee Px)$   
 $\therefore (x)(Mx \supset Px)$

## Inductive Logic I: Arguments from Analogy and Causal Reasoning

### I. Inductive Logics

Back in Chapter 1, we made a distinction between deductive and inductive arguments. While deductive arguments attempt to provide premises that *guarantee* their conclusions, inductive arguments are less ambitious. They merely aim to provide premises that make the conclusion more probable. Because of this difference, it is inappropriate to evaluate deductive and inductive arguments by the same standards. We do not use the terms ‘valid’ and ‘invalid’ when evaluating inductive arguments. Technically, they’re all invalid because their premises don’t guarantee their conclusions; but that’s not a fair evaluation, since inductive arguments don’t even pretend to try to provide such a guarantee. Rather, we say of inductive arguments that they are strong or weak, cogent or incogent — the more probable the conclusion in light of the premises, the stronger the inductive argument; the less probable the conclusion, the weaker. These judgments can change in light of new information. Additional evidence may have the effect of making the conclusion more or less probable—of strengthening or weakening the argument.

The topic of this chapter and the next will be inductive logic: we will be learning about the various types of inductive arguments and how to evaluate them. Inductive arguments are a rather motley bunch. They come in a wide variety of forms that can vary according to subject matter; they resist the uniform treatment we were able to provide for their deductive cousins. We will have to examine a wide variety of approaches—different inductive methods. While all inductive arguments have in common that they attempt to give their conclusions more probable, it is not always possible for us to make precise judgments about exactly *how* probable their conclusions are in light of their premises. When that is the case, we will make *relative* judgments: this argument is stronger or weaker than that argument, though I can’t say how much stronger or weaker, precisely. Sometimes, however, it will be possible to render precise judgments about the probability of conclusions. In order for us to apply inferences to statistics and probabilities, Chapter 8 will require us to acquire basic skills in calculating probabilities. With those in hand, we will be in a position to model an ideally rational approach to revising our judgments about the strength of inductive arguments in light of new evidence. In addition, since so many inductive arguments use statistics, it will be necessary for us to acquire a basic understanding of some fundamental statistical concepts. With these in hand, we will be in a position to recognize the most common types of statistical fallacies—mistakes and intentionally misleading arguments that use statistics to lead us astray.

Probability and statistics will be the subject of the next chapter. In this chapter, we will look at two very common types of inductive reasoning: arguments from analogy and inferences involving causation. The former are quite common in everyday life; the latter are the primary methods of scientific and medical research. Each type of reasoning exhibits certain patterns, and we will look at the general forms analogical and causal arguments; we want to develop the skill of recognizing how particular instances of reasoning fit these general patterns. We will also learn how these types of arguments are evaluated. For arguments from analogy, we will identify the criteria that we use to make relative judgments about strength and weakness. For causal reasoning, we will compare the various forms of inference to identify those most likely to produce reliable results, and we will examine some of the pitfalls peculiar to each that can lead to errors.

## II. Arguments from Analogy

Analogical reasoning is ubiquitous in everyday life. We rely on analogies—similarities between present circumstances and those we've already experienced—to guide our actions. We use comparisons to familiar people, places, and things to guide our evaluations of novel ones. We criticize people's arguments based on their resemblance to obviously absurd lines of reasoning.

In this section, we will look at the various uses of analogical reasoning. Along the way, we will identify a general pattern that all arguments from analogy follow and learn how to show that particular arguments fit the pattern. We will then turn to the evaluation of analogical arguments: we will identify six criteria that govern our judgments about the relative strength of these arguments. Finally, we will look at the use of analogies to refute other arguments.

### *The Form of Analogical Arguments*

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

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

Notice, first, that this is an inductive argument. The conclusion, that I will like *The Wolf of Wall Street* is not guaranteed by the premises; as a matter of fact, my prediction was wrong and I really didn't care for the film. But our real focus here is on the fact that the prediction was made on the basis of an analogy. Actually, several analogies, between *The Wolf of Wall Street*, on the one hand, and all the other Scorsese/DiCaprio collaborations on the other. The new film is *similar* in important respects to the older ones; I liked all of those; so, I'll probably like the new one.

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

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

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

Analogical arguments all share the same basic structure. A set of examples all share a bunch of properties as well as sharing the particular feature we looking to extrapolate onto our subject. Our subject shows the same initial properties and so we infer that the pattern will continue and that we can infer the unknown feature onto our subject as well. We can lay out this form schematically as follows:

$a_1, a_2, \dots, a_n$ , and  $c$  all have  $P_1, P_2, \dots, P_k$   
 $a_1, a_2, \dots, a_n$  all have  $Q$   
 $\therefore c$  has  $Q$

This is an abstract schema, and it's going to take some getting used to, but it represents the form of analogical reasoning succinctly and clearly. Arguments from analogy have two premises and a conclusion. The first premise establishes an analogy. The analogy is between some thing, marked 'c' in the schema, and some number of other things, marked 'a<sub>1</sub>', 'a<sub>2</sub>', and so on in the schema. We can refer to these as the "analogues". They're the things that are similar, analogous to c. This schema is meant to cover every possible argument from analogy, so we do not specify a particular number of analogues; the last one on the list is marked 'a<sub>n</sub>', where 'n' is a variable standing for any number whatsoever. There may be only one analogue; there may be a hundred. What's important is that the analogues are similar to the thing designated by 'c'. What makes different things similar? They have stuff in common; they share properties. Those properties—the similarities between the analogues and c—are marked 'P<sub>1</sub>', 'P<sub>2</sub>', and so on in the diagram. Again, we don't specify a particular number of properties shared: the last is marked 'P<sub>k</sub>', where 'k' is just another variable (we don't use 'n' again, because the number

of analogues and the number of properties can of course be different). This is because our schema is generic: every argument from analogy fits into the framework; there may be any number of properties involved in any particular argument. Anyway, the first premise establishes the analogy: *c* and the analogues are similar because they have various things in common— $P_1, P_2, P_3, \dots, P_k$ .

Notice that '*c*' is missing from the second premise. The second premise only concerns the analogues: it says that they have some property in common, designated '*Q*' to highlight the fact that it's not among the properties listed in the first premise. It's a separate property. It's the very property we're trying to establish, in the conclusion, that *c* has ('*c*' is for conclusion). The thinking is something like this: *c* and the analogues are similar in so many ways (first premise); the analogues have this additional thing in common (*Q* in the second premise); so, *c* is probably like that, too (conclusion: *c* has *Q*).

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

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

The conclusion is something like 'I will like *The Wolf of Wall Street*'. Putting it that way, and looking at the general form of the conclusion of analogical arguments (*c* has *Q*), it's tempting to say that '*c*' designates me, while the property *Q* is something like 'liking *The Wolf of Wall Street*'. But that's not right. The thing that '*c*' designates has to be involved in the analogy in the first premise; it has to be the thing that's similar to the analogues. The analogy that this argument hinges on is between the various *movies*. It's not I that '*c*' corresponds to; it's the movie we're making the prediction about. *The Wolf of Wall Street* is what '*c*' picks out. What property are we predicting it will have? Something like 'liked by me'. The analogues, the *a*'s in the schema, are the other movies: *Gangs of New York*, *The Aviator*, *The Departed*, and *Shutter Island*. (In this example, *n* is 4; the movies are *a*<sub>1</sub>, *a*<sub>2</sub>, *a*<sub>3</sub>, and *a*<sub>4</sub>.) These we know have the property *Q* (liked by me): I had already seen and liked these movies. That's the second premise: that the analogues have *Q*. Finally, the first premise, which establishes the analogy among all the movies. What do they have in common? They were all directed by Martin Scorsese, and they all starred Leonardo DiCaprio.

Those are the *P*'s—the properties they all share. *P*<sub>1</sub> is 'directed by Scorsese'; *P*<sub>2</sub> is 'stars DiCaprio'.

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

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

Again, looking at the conclusion—'Eating pork is immoral'—and looking at the general form of conclusions for analogical arguments—'*c* has *Q*'—it's tempting to just read off from the syntax of the sentence that '*c*' stands for 'eating pork' and *Q* for 'is immoral'. But that's not right. Focus on the analogy: what things are being compared to one another? It's the animals: pigs, dogs, and dolphins; those are our *a*'s and *c*. To determine which one is picked out by '*c*', we ask which animal is involved in the conclusion. It's pigs; they are picked out by '*c*'. So we have to paraphrase our conclusion so that it fits the form '*c* has *Q*', where '*c*' stands for pigs. Something like 'Pigs shouldn't be eaten' would work. So *Q* is the property 'shouldn't be eaten'. The analogues are dogs and dolphins. They clearly have the property: as the argument notes, (most) everybody agrees they shouldn't be eaten. This is the second premise. And the first establishes the analogy. What do pigs have in common with dogs and dolphins? They're smart, cute, and playful. *P*<sub>1</sub> = 'is smart'; *P*<sub>2</sub> = 'is cute'; and *P*<sub>3</sub> = 'is playful'.

## *The Evaluation of Analogical Arguments*

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

a *metacognitive perspective* on the practice of evaluating arguments from analogy; we want to think about a type of thinking that we typically engage in without much conscious deliberation. We want to identify the criteria that we rely on to evaluate analogical reasoning—criteria that we apply without necessarily realizing that we’re applying them. Achieving such metacognitive awareness is useful insofar as it makes us more self-aware, critical, and therefore effective reasoners.

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

As we mentioned, while it’s often impossible to evaluate inductive arguments by giving a precise probability of its conclusion, it is possible to make *relative* judgments about strength and weakness. Recall, new information can change the probability of the conclusion of an inductive argument. We can make relative judgments of like this: if we add this new information as a premise, the new argument is stronger/weaker than the old argument; that is, the new information makes the conclusion more/less likely.

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

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

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

This is a simple argument from analogy, in which the future is predicted based on past experience. It fits the schema for analogical arguments: the new course she has signed up for is designated by

‘c’; the property we’re predicting it has (Q) is that it is a course Alice will get an A in; the analogues are the four previous courses she’s taken; what they have in common with the new course (P<sub>1</sub>) is that they are also Philosophy classes; and they all have the property Q—Sally got an A in each.

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

## Number of Analogues

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

It’s stronger! We’ve got Alice getting an A 14 times in a row instead of only four. That clearly makes the conclusion more probable. (How much more? Again, it doesn’t matter.)

What we did in this case is add more analogues. This reveals a general rule: other things being equal, the more analogues in an analogical argument, the stronger the argument (and conversely, the fewer analogues, the weaker). The number of analogues is one of the criteria we use to evaluate arguments from analogy.



## Variety of Analogues

You'll notice that the original argument doesn't give us much information about the four courses Alice succeeded in previously and the new course she's about to take. All we know is that they're all Philosophy courses. Suppose we tweak things. We're still in the dark about the new course Alice is about to take, but we know a bit more about the other four: one was a course in Ancient Greek Philosophy; one was a course on Contemporary Ethical Theories; one was a course in Formal Logic; and the last one was a course in the Philosophy of Mind. Given this new information, are we more or less confident that she will succeed in the new course, whose topic is unknown to us? Is the argument stronger or weaker than the baseline argument?

It is stronger. We don't know what kind of Philosophy course Alice is about to take, but this new information gives us an indication that it doesn't really matter. She was able to succeed in a wide variety of courses, from Mind to Logic, from Ancient Greek to Contemporary Ethics. This is evidence that Alice is good at Philosophy generally, so that no matter what kind of course she's about to take, she'll probably do well in it.

Again, this points to a general principle about how we evaluate analogical arguments: other things being equal, the more variety there is among the analogues, the stronger the argument (and conversely, the less variety, the weaker).

## Number of Similarities

In the baseline argument, the only thing the four previous courses and the new course have in common is that they're Philosophy classes. Suppose we change that. Our newly tweaked argument predicts that Alice will get an A in the new course, which, like the four she succeeded in before, is cross-listed in the Department of Religious Studies and covers topics in the Philosophy of Religion. Given this new information—that the new course and the four older courses were similar in ways we weren't aware of—are we more or less confident in the prediction that Alice will get another A? Is the argument stronger or weaker than the baseline argument?

Again, it is stronger. Unlike the last example, this tweak gives us new information both about the four previous courses *and* the new one. The upshot of that information is that they're even more similar than we knew; that is, they have more properties in common. To  $P_1$  = 'is a Philosophy course' we can add  $P_2$  = 'is cross-listed with Religious Studies' and  $P_3$  = 'covers topics in Philosophy of Religion'. The more properties things have in common, the stronger the analogy between them. The stronger the analogy, the stronger the argument based on that analogy. We now know not just that Alice did well in not just in Philosophy classes—but specifically in classes covering the Philosophy of Religion; *and* we know that the new class she's taking is *also* a Philosophy of Religion class. I'm much more confident predicting she'll do well again than I was when all I knew was that all the classes were Philosophy; the new one could've been in a different topic that she wouldn't have liked.

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

## Number of Differences

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

Suppose we add new information to our baseline argument: the four Philosophy courses Alice did well in before were all courses in the Philosophy of Mind; the new course is about the history of Ancient Greek Philosophy. Given this new information, are we more or less confident that she will succeed in the new course? Is the argument stronger or weaker than the baseline argument?

Clearly, the argument is weaker. The new course is on a completely different topic than the other ones. She did well in four straight Philosophy of Mind courses, but Ancient Greek Philosophy is quite different. I'm less confident that she'll get an A than I was before.

If I add more differences, the argument gets even weaker. Supposing the four Philosophy of Mind courses were all taught by the same professor (the person in the department whose expertise is in that area), but the Ancient Greek Philosophy course is taught by someone different (the department's specialist in that topic). Different subject matter, different teachers: I'm even less optimistic about Alice's continued success.

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

## Relevance of Similarities and Differences

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

Similarly, if we tweak the original argument to add a difference between the new class and the other four, to the effect that while all of the four older classes were in the same building, while the new one is in a different building, there is no effect on our confidence in the conclusion. Again, the building in which a class meets is simply not relevant to how well someone does.

Contrast these cases with the new information that the new course and the previous four are all taught by the same professor. Now that strengthens the argument! Alice has gotten an A four times in a row from this professor—all the more reason to expect she'll receive another one. This tidbit strengthens the argument because the new similarity—the same person teaches all the courses—is relevant to the prediction we're making—that Alice will do well. Who teaches a class can make a difference to how students do—either because they're easy graders, or because they're great teachers, or because the student and the teacher are in tune with one another, etc. Even a difference between the analogues and the item in the conclusion, with the right kind of relevance, can strengthen an argument. Suppose the other four philosophy classes were taught by the same teacher, but the new one is taught by a TA—who just happens to be her boyfriend. That's a difference, but one that makes the conclusion—that Alice will do well—more probable.

Generally speaking, careful attention must be paid to the relevance of any similarities and differences to the property in the conclusion; the effect on strength varies.

## Modesty/Ambition of the Conclusion

Suppose we leave everything about the premises in the original baseline argument the same: four Philosophy classes, an A in each, new Philosophy class. Instead of adding to that part of the argument, we'll tweak the conclusion. Instead of predicting that Alice will get an A in the class, we'll predict that she'll pass the course. Are we more or less confident that this prediction will come true? Is the new, tweaked argument stronger or weaker than the baseline argument?

It's stronger. We are more confident in the prediction that Alice will pass than we are in the prediction that she will get another A, for the simple reason that it's much easier to pass than it is to get an A. That is, the prediction of passing is a much more *modest* prediction than the prediction of an A.

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

General principle: the more modest the argument’s conclusion, the stronger the argument; the more ambitious, the weaker.

## *Refutation by Analogy*

We can use arguments from analogy for a specific logical task: refuting someone else’s argument, showing that it’s bad. Recall the case of deductive arguments. To refute those—to show that they are bad, i.e., invalid—we had to produce a counterexample—a new argument with the same logical form as the original that was obviously invalid, in that its premises were in fact true and its conclusion in fact false. We can use a similar procedure to refute inductive arguments. Of course, the standard of evaluation is different for induction: we don’t judge them according to the black and white standard of validity. And as a result, our judgments have less to do with form than with content. Nevertheless, refutation along similar lines is possible, and analogies are the key to the technique.

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

Consider the following:

“Duck Dynasty” star and Duck Commander CEO Willie Robertson said he supports Trump because both of them have been successful businessmen and stars of reality TV shows.

By that logic, does that mean Hugh Hefner’s success with “Playboy” and his occasional appearances on “Bad Girls Club” warrant him as a worthy president? Actually, I’d still be more likely to vote for Hefner than Trump.<sup>1</sup>

The author is refuting the argument of Willie Robertson, the “Duck Dynasty” star. Robertson’s argument is something like this: Trump is a successful businessman and reality TV star; therefore, he would be a good president. To refute this, the author produces an analogous argument—Hugh Hefner is a successful businessman and reality TV star; therefore, Hugh Hefner would make a good president. And our refuter thinks that second argument is obviously bad. What makes it obviously bad is that it has a conclusion that nobody would agree with: that Hugh Hefner would make a good president. That’s how these refutations work. They attempt to demonstrate that the original argument is lousy by showing that you can use the same or very similar reasoning to arrive at an absurd conclusion.

Here’s another example, from a group called “Iowans for Public Education”. Next to a picture of an apparently well-to-do lady is the following text:

“My husband and I have decided the local parks just aren’t good enough for our kids. We’d rather use the country club, and we are hoping state tax dollars will pay for it. We are advocating for Park Savings Accounts,

1. Austin Faulds, “Weird celebrity endorsements fit for weird election,” *Indiana Daily Student*, 10/12/16, <http://www.idsnews.com/article/2016/10/weird-celebrity-endorsements-for-weird-election>.

or PSAs. We promise to no longer use the local parks. To hell with anyone else or the community as a whole. We want our tax dollars to be used to make the best choice for our family.”

Sound ridiculous? Tell your legislator to vote NO on Education Savings Accounts (ESAs), aka school vouchers.

The argument that Iowans for Public Education put in the mouth of the lady on the poster is meant to refute reasoning used by advocates for “school choice”, who say that they ought to have the right to opt out of public education and keep the tax dollars they would otherwise pay for public schools and use it to pay to send their kids to private schools. A similar line of reasoning sounds pretty crazy when you replace public schools with public parks and private schools with country clubs.

Since these sorts of refutations rely on analogies, they are only as strong as the analogy between the refuting and refuted arguments. There is room for dispute on that question. Advocates for school vouchers might point out that schools and parks are completely different things, that schools are much more important to the future prospects of children, and that given the importance of education, families should have to right choose what they think is best. Or something like that. The point is, the kinds of knock-down refutations that were possible for deductive arguments are not possible for inductive arguments. There is always room for further debate.

## EXERCISES

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

$a_1, a_2, \dots, a_n$ , and  $c$  all have  $P_1, P_2, \dots, P_k$

$a_2, \dots, a_n$  all have  $Q$

$\therefore c$  has  $Q$

(a) You should really eat at Papa Giorgio’s; you’ll love it. It’s just like Mama DiSilvio’s and Matteo’s, which I know you love: they serve old-fashioned Italian-American food, they have a laid-back atmosphere, and the wine list is extensive.

(b) George R.R. Martin deserves to rank among the greats in the fantasy literature genre. Like C.S. Lewis and J.R.R. Tolkien before him, he has created a richly detailed world, populated it with compelling characters, and told a tale that is not only exciting, but which features universal and timeless themes concerning human nature.

(c) Yes, African Americans are incarcerated at higher rates than whites. But blaming this on systemic racial bias in the criminal justice system is absurd. That’s like saying the NBA is racist because there are more black players than white players, or claiming that the medical establishment is racist because African Americans die young more often.

2. Consider the following base-line argument:

I’ve taken vacations to Florida six times before, and I’ve enjoyed each visit. I’m planning to go to Florida again this year, and I fully expect yet another enjoyable vacation.

Decide whether each of the following changes produces an argument that’s weaker or stronger than the baseline argument, and indicate which of the six criteria for evaluating analogical arguments justifies that judgment.

(a) All of my trips were visits to Disney World, and this one will be no different.

(b) In fact, I’ve vacationed in Florida 60 times and enjoyed every visit.

(c) I expect that I will enjoy this trip so much I will decide to *move* to Florida.

(d) On my previous visits to Florida, I’ve gone to the beaches, the theme parks, the Everglades National Park, and various cities, from Jacksonville to Key West.

(e) I’ve always flown to Florida on Delta Airlines in the past; this time I’m going on a United flight.

(f) All of my past visits were during the winter months; this time I’m going in the summer.

(g) I predict that I will find this trip more enjoyable than a visit to the dentist.

- (h) I've only been to Florida once before.
- (i) On my previous visits, I drove to Florida in my Dodge minivan, and I'm planning on driving the van down again this time.
- (j) All my visits have been to Daytona Beach for the Daytona 500; same thing this time.
- (k) I've stayed in beachside bungalows, big fancy hotels, time-share condominiums—even a shack out in the swamp.

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

(a) Republicans tell us that, because at some point 40 years from now a shortfall in revenue for Social Security is projected, we should cut benefits now. Cut them now because we might have to cut them in the future? I've got a medium-sized tree in my yard. 40 years from now, it may grow so large that its branches hang over my roof. Should I chop it down?

(b) Opponents of gay marriage tell us that it flies in the face of a tradition going back millennia, that marriage is between a man and a woman. There were lots of traditions that lasted a long time: the tradition that it was OK for some people to own other people as slaves, the tradition that women couldn't participate in the electoral process—the list goes on. That it's traditional doesn't make it right.

(c) Some people claim that their children should be exempted from getting vaccinated for common diseases because the practice conflicts with their religious beliefs. But religion can't be used to justify just anything. If a priest tried to defend himself against charges of abusing children by claiming that such practices were a form of religious expression, would we let him get away with it?

### III. Causal Reasoning

Inductive arguments are used to support claims about cause and effect. These arguments come in a number of different forms. The most straightforward is what is called *enumerative induction*. This is an argument that makes a (non-hasty) generalization, inferring that one event or type of event causes another on the basis of a (large) number of particular observations of the cause immediately preceding the effect. To use a very famous example (from the history of philosophy, due to David Hume, the 18<sup>th</sup> century Scottish philosopher who had much to say about cause and effect and inductive reasoning), we can infer from observations of a number of billiard-ball collisions that the first ball colliding with the second causes the second ball to move. Or we can infer from a number of observations of drunkenness following the consumption of alcoholic beverages that imbibing alcohol causes one to become drunk.

This is all well and good, so far as it goes.<sup>2</sup> It just doesn't go very far. If we want to establish a robust knowledge of what causes the natural phenomena we're interested in, we need techniques that are more sophisticated than simple enumerative induction. There are such techniques. These are patterns of reasoning identified and catalogued by the 19<sup>th</sup> century English philosopher, scientist, logician, and politician John Stuart Mill. The inferential forms Mill enumerated have come to be called "Mill's Methods", because he thought of them as tools to be used in the investigation of nature—methods of discovering the causes of natural phenomena. In this section, we will look at Mill's Methods each in turn (there are five of them), using examples to illustrate each. We will finish with a discussion of the limitations of the methods and the difficulty of isolating causes.

2. Setting aside Hume's philosophical skepticism about our ability to know that one thing causes another and about the conclusiveness of inductive reasoning.

## The Meaning(s) of ‘Cause’

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

We're not going to wade into those philosophical waters; they're too deep. Instead, we'll merely dip our toes in, by making a preliminary observation about the word 'cause'—an observation that gives some hint as to why it's been the subject of so much philosophical deliberation for so long. The observation is this: there are a number of distinct, but perfectly acceptable ways that we use the word 'cause' in everyday language. We attach different incompatible meanings to the term in different contexts.

Consider this scenario: I'm in my backyard vegetable garden with my younger daughter (age 4 at the time). She's "helping" me in my labors by watering some of the plants. She asks, "Daddy, why do we have to water the plants?" I might reply, "We do that because water causes the plants to grow." This is a perfectly ordinary claim about cause and effect; it is uncontroversial and true. What do I mean by 'causes' in this sentence? I mean that water is a *necessary condition* for the plants to grow. Without water, there will be no growth. It is not a sufficient condition for plant-growth, though: you also need sunlight, good soil, etc.

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

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

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

## Mill's Methods

John Stuart Mill identified five different patterns of reasoning that one could use to discover causes. These are argument forms, the conclusions of which involve a claim to the effect that one thing causes (or is causally related to) another. They can be used alone or in combination, depending on the circumstances. As was the case with analogical reasoning, these are patterns of inference that we already employ unreflectively in everyday life. The benefit in making them explicit and subjecting them to critical scrutiny is that we thereby achieve a metacognitive perspective—a perspective from which we can become more self-aware, effective reasoners. This is especially important in the context of causal reasoning, since, as we shall see, there are many pitfalls in this domain that we are prone to fall into, many common errors that people make when thinking about cause and effect.

## Method of Agreement

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

This little story is an instance of Mill's Method of Agreement. It's a pattern of reasoning that one can use to figure out the cause of some phenomenon of interest. In this case, the phenomenon I want to discover the cause of is my recent episodes of heartburn. I eventually figure out that the cause is Mountain Dew. We could sum up the reasoning pattern abstractly thus:

We want to find the cause of a phenomenon, call it X. We examine a variety of circumstances in which X occurs, looking for potential causes. The circumstances differ in various ways, but they each have in common that they feature the same potential cause, call it A. We conclude that A causes X.

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

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

What causal factor is present whenever the phenomenon of interest is present?

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

## Method of Difference

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

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

We want to find the cause of a phenomenon, call it X. We examine a variety of circumstances in which X occurs, looking for potential causes. The circumstances differ in various ways, but they each have in common that when we remove from them a potential cause—call it A—the phenomenon disappears. We conclude that A causes X.

If we introduce the same *difference* in all of the circumstances—removing the causal factor—we see the same effect—disappearance of the phenomenon. Hence, the Method of Difference. In our story, the phenomenon we wanted

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

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

What causal factor is absent whenever the phenomenon of interest is absent?

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

## Joint Method of Agreement and Difference

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

The Joint Method is an important tool in medical research. It's the pattern of reasoning used in what we call *controlled studies*. In such a study, we split our subjects into two groups, one of which is the "control" group. An example shows how this works. Suppose I've formulated a pill that I think is a miracle cure for baldness. I'm gonna be rich! But first, I need to see if it really works. So I gather a bunch of bald men together for a controlled study. One group gets the actual drug; the other, control group, gets a sugar pill—not the real drug at all, but a mere placebo. Then I wait and see what happens. If my drug is a good as I think it is, two things will happen: first, the group that got the drug will grow new hair; and second, the group that got the placebo *won't* grow new hair. If either of these things fails to happen, it's back to the drawing board. Obviously, if the group that got the drug didn't get any new hair, my baldness cure is a dud. But in addition, if the group that got the mere placebo grew new hair, then something else besides my drug has to be the cause.

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

## Method of Residues

I'm running a business. Let's call it LogiCorp. For a modest fee, the highly trained logicians at LogiCorp will evaluate all of your deductive arguments, issuing Certificates of Validity (or Invalidity) that are legally binding in all fifty states. Satisfaction guaranteed. Anyway, as should be obvious from that brief description of the business model, LogiCorp is a highly profitable enterprise. But last year's results were disappointing. Profits were down 20% from the year before. Some of this was expected. We undertook a renovation of the LogiCorp World Headquarters that year, and the cost had an effect on our bottom line: half of the lost profits, 10%, can be chalked up to the renovation expenses. Also, as healthcare costs continue to rise, we had to spend additional money on our employees' benefits packages; these expenditures account for an additional 3% of profit shortfall. Finally, another portion of the drop in profits can be explained by the entry of a competitor into the marketplace. The upstart firm Arguments R Us, with its fast turnaround times and ultra-cheap prices, has been cutting into our market share. Their services are totally inferior to ours (you should see the shoddy shading technique in their Venn Diagrams!) and LogiCorp will crush them eventually, but for now they're hurting our business: competition from Arguments R Us accounts for a 5% drop in our profits.

As CEO, I was of course aware of all these potential problems throughout the year, so when I looked at the numbers at the end, I wasn't surprised. But, when I added up the contributions from the three factors I knew about—10% from



the renovation, 3% from the healthcare expenditures, 5% from outside competition—I came up short. Those causes only account for an 18% shortfall in profit, but we were down 20% on the year; there was an extra 2% shortfall that I couldn't explain. I'm a suspicious guy, so I hired an outside security firm to monitor the activities of various highly placed employees at my firm. And I'm glad I did! Turns out my Chief Financial Officer had been taking lavish weekend vacations to Las Vegas and charging his expenses to the company credit card. His thievery surely accounts for the extra 2%. I immediately fired the jerk. (Maybe he can get a job with Arguments R Us.)

This little story presents an instance of Mill's Method of Residues. 'Residue' in this context just means the remainder, that which is left over. The pattern of reasoning, put abstractly, runs something like this:

We observe a series of phenomena, call them  $X_1, X_2, X_3, \dots, X_n$ . As a matter of background knowledge, we know that  $X_1$  is caused by  $A_1$ , that  $X_2$  is caused by  $A_2$ , and so on. But when we exhaust our background knowledge of the causes of phenomena, we're left with one,  $X_n$ , that is inexplicable in those terms. So we must seek out an additional causal factor,  $A_n$ , as the cause of  $X_n$ .

The leftover phenomenon,  $X_n$ , inexplicable in terms of our background knowledge, is the residue. In our story, that was the additional 2% profit shortfall that couldn't be explained in terms of the causal factors we were already aware of, namely the headquarters renovation ( $A_1$ , which caused  $X_1$ , a 10% shortfall), the healthcare expenses ( $A_2$ , which caused  $X_2$ , a 3% shortfall), and the competition from Arguments R Us ( $A_3$ , which caused  $X_3$ , a 5% shortfall). We had to search for another, previously unknown cause for the final, residual 2%.

## Method of Concomitant Variation

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

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

(These are revolutionary insights, I know. My next get-rich-quick scheme is to popularize one of those fad diets. Instead of recommending eating nothing but bacon or drinking nothing but smoothies made of kale and yogurt, my fad diet will be the "Eat Less, Move More" plan. I'm gonna be rich!)

I know about the cause-and-effect relationships above because of the Method of Concomitant Variation. Put abstractly, this pattern of reasoning goes something like this:

We observe that, holding other factors constant, an increase or decrease in some causal factor  $A$  is always accompanied by a corresponding increase or decrease in some phenomenon  $X$ . We conclude that  $A$  and  $X$  are causally related.

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

In our first example, calorie consumption ( $A$ ) and weight ( $X$ ) vary directly. As calorie consumption increases, weight increases; and as calorie consumption decreases, weight decreases. In our second example, exercise ( $A$ ) and weight ( $X$ ) vary inversely. As exercise increases, weight decreases; and as exercise decreases, weight increases.

Either way, when things change together in this way, when they vary concomitantly, we conclude that they are causally related.

## The Difficulty of Isolating Causes

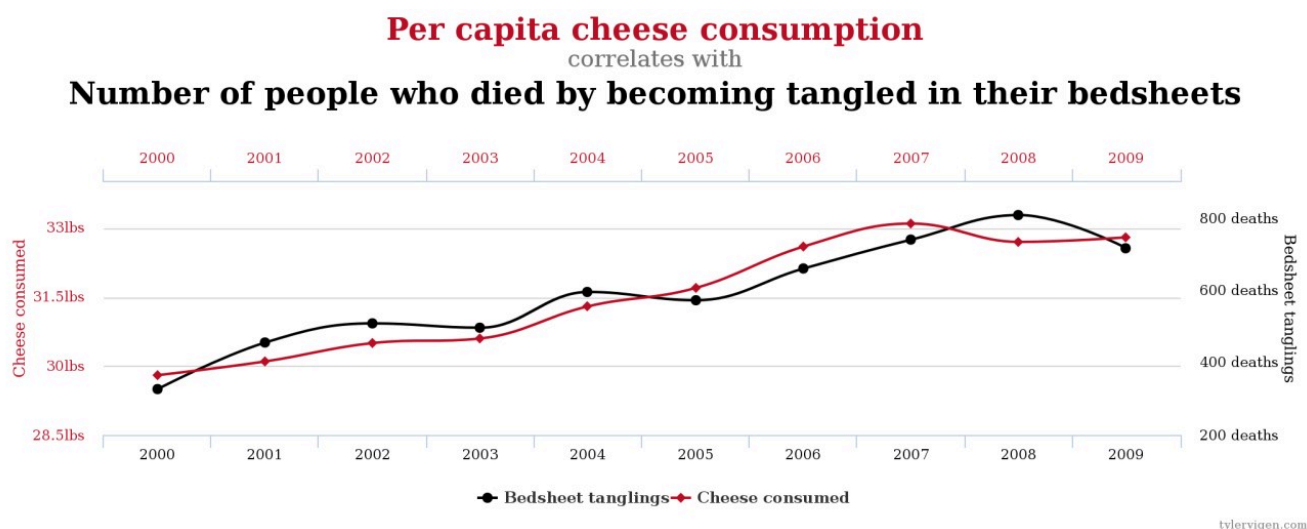
Mill's Methods are useful in discovering the causes of phenomena in the world, but their usefulness should not be overstated. Unless they are employed thoughtfully, they can lead an investigator astray. A classic example of this is the parable of the drunken logician.<sup>3</sup> After a long day at work on a Monday, a certain logician heads home wanting to unwind. So he mixes himself a "7 and 7"—Seagram's 7, Crown whiskey, and 7-Up. It tastes so good, he makes another—and another, and another. He drinks seven of these cocktails, passes out in his clothes, and wakes up feeling terrible (headache, nausea, etc.). On Tuesday, after dragging himself into work, toughing it through the day, then finally getting home, he decides to take the edge off with a different drink: brandy and 7-Up. He gets carried away again, and ends up drinking seven of these cocktails, with the same result: passing out in his clothes and waking up feeling awful on Wednesday. So, on Wednesday night, our logician decides to mix things up again: scotch and 7-Up. He drinks seven of these; same results. But he perseveres: Thursday night, it's seven vodka and 7-Ups; another blistering hangover on Friday. So on Friday at work, he sits down to figure out what's going on. He's got a phenomenon—hangover symptoms every morning of that week—that he wants to discover the cause of. He's a professional logician, intimately familiar with Mill's Methods, so he figures he ought to be able to discover the cause. He looks back at the week and uses the Method of Agreement, asking, "What factor was present every time the phenomenon was?" He concludes that the cause of his hangovers is 7-Up.

Our drunken logician applied the Method of Agreement correctly: 7-Up was indeed present every time. But it clearly wasn't the cause of his hangovers. The lesson is that Mill's Methods are useful tools for discovering causes, but their results are not always definitive. Uncritical application of the methods can lead one astray. This is especially true of the Method of Concomitant Variation. You may have heard the old saw that "correlation does not imply causation." It's useful to keep this corrective in mind when using the Method of Concomitant Variation. That two things vary concomitantly is a hint that they may be causally related, but it is not definitive proof that they are. They may be separate effects of a different, unknown cause; they may be completely causally unrelated. It is true, for example, that among children, shoe size and reading ability vary directly: children with bigger feet are better readers than those with smaller feet. Wow! So large feet cause better reading? Of course not. Larger feet and better reading ability are both effects of the same cause: getting older. Older kids wear bigger shoes than younger kids, and they also do better on reading tests. Duh. It is also true, for example, that hospital quality and death rate vary directly: that is, the higher quality the hospital (prestige of doctors, training of staff, sophistication of equipment, etc.), on average, the higher the death rate at that hospital. That's counterintuitive! Does that mean that high hospital quality causes high death rates? Of course not. Better hospitals have higher mortality rates because the extremely sick, most badly injured patients are taken to those hospitals, rather than the ones with lower-quality staff and equipment. Alas, these people die more often, but not because they're at a good hospital; it's exactly the reverse.

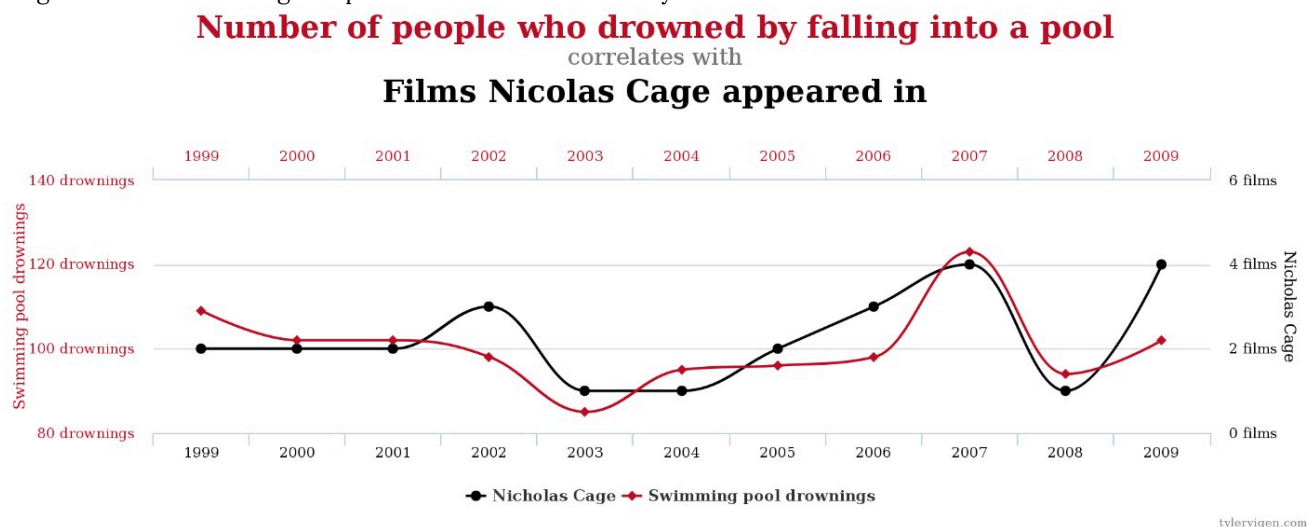
Spurious correlations—those that don't involve any causal connection at all—are easy to find in the age of "big data." With publicly available databases archiving large amounts of data, and computers with the processing power to search them and look for correlations, it is possible to find many examples of phenomena that vary concomitantly but are obviously not causally connected. A very clever person named Tyler Vigen set about doing this and created a website where he posted his (often very amusing) discoveries.<sup>4</sup> For example, he found that between 2000 and 2009, per capita cheese consumption among Americans was very closely correlated with the number of deaths caused by people becoming entangled in their bedsheets:

3. Inspired by Copi and Cohen, p. 547.

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



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



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

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

in treating the common cold.<sup>5</sup> We gather 100 subjects experiencing the onset of cold symptoms. We want one group of 50 to get vitamin C supplements, and one group of 50—the control group—not to receive them. How do we decide who gets placed into which group? We could ask for volunteers. But doing so might create a systematic difference between the two groups. People who hear “vitamin C” and think, “yeah, that’s the group for me” might be people who are more inclined to eat fruits and vegetables, for example, and might therefore be healthier on average than people who are turned off by the idea of receiving vitamin C supplements. This difference between the groups might lead to different results between the how their colds progress. Instead of asking for volunteers, we might just assign the first 50 people who show up to the vitamin C group, and the last 50 to the control group. But this could lead to differences, as well. The people who show up earlier might be early-risers, who might be healthier on average than those who straggle in late.

The best way to avoid systematic differences between test and control groups is to randomly assign subjects to each. We refer to studies conducted this way as *randomized controlled studies*. And besides randomization, other measures can be taken to improve reliability. The best kinds of controlled studies are “double-blind”. This means that neither the subjects nor the people conducting the study know which group is the control and which group is receiving the actual treatment. (This information is hidden from the researchers only while the study is ongoing; they are told later, of course, so they can interpret the results.) This measure is necessary because of the psychological tendency for people’s observations to be biased based on their expectations. For example, if the control group in our vitamin C experiment knew they were not getting any treatment for their colds, they might be more inclined to report that they weren’t feeling any better. Conversely, if the members of the group receiving the vitamin supplements knew that they were getting treated, they might be more inclined to report that their symptoms weren’t as bad. This is why the usual practice is to keep subjects in the dark about which group they’re in, giving a placebo to the members of the control group. It’s important to keep the people conducting the study “blind” for the same reasons. If they knew which group was which, they might be more inclined to observe improvement in the test group and a lack of improvement in the control group. In addition, in their interactions with the subjects, they may unknowingly give away information about which group was which via subconscious signals.

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

## EXERCISES

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

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

Alfonse, Bertram, Claire, Dominic, Ernesto, and Francine all go out to dinner at a local greasy spoon. There are

5. Despite widespread belief that it is, researchers have found very little evidence to support this claim.

six items on the menu: shrimp cocktail, mushroom/barley soup, burger, fries, steamed carrots, and ice cream. This is what they ate:

Alfonse: shrimp, soup, fries

Bertram: burger, fries, carrots, ice cream

Claire: soup, burger, fries, carrots

Dominic: shrimp, soup, fries, ice cream

Ernesto: burger, fries, carrots

Francine: ice cream

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

(a) Using only the Method of Agreement, how far can we narrow down the list of possible causes for the food poisoning?

(b) Using only the Method of Difference, how far can we narrow down the list of possible causes for the food poisoning?

(c) Using the Joint Method, we can identify the cause. What is it?

3. For each of the following, identify which of Mill's Methods is being used to draw the causal conclusion.

(a) A farmer noticed a marked increase in crop yields for the season. He started using a new and improved fertilizer that year, and the weather was particularly ideal—just enough rain and sunshine. Nevertheless, the increase was greater than could be explained by these factors. So he looked into it and discovered that his fields had been colonized by hedgehogs, who prey on the kinds of insect pests that usually eat crops.

(b) I've been looking for ways to improve the flavor of my vegan chili. I read on a website that adding soy sauce can help: it has lots of umami flavor, and that can help compensate for the lack of meat. So the other day, I made two batches of my chili, one using my usual recipe, and the other made exactly the same way, except for the addition of soy sauce. I invited a bunch of friends over for a blind taste test, and sure enough, the chili with the soy sauce was the overwhelming favorite!

(c) The mere presence of guns in circulation can lead to higher murder rates. The data are clear on this. In countries with higher numbers of guns per capita, the murder rate is higher; and in countries with lower numbers of guns per capita, the murder rate is correspondingly lower.

(d) There's a simple way to end mass shootings: outlaw semiautomatic weapons. In 1996, Australia suffered the worst mass shooting episode in its history, when a man in Tasmania used two semiautomatic rifles to kill 35 people (and wound an additional 19). The Australian government responded by making such weapons illegal. There hasn't been a mass shooting in Australia since.

(e) A pediatric oncologist was faced with a number of cases of childhood leukemia over a short period of time. Puzzled, he conducted thorough examinations of all the children, and also compared their living situations. He was surprised to discover that all of the children lived in houses that were located very close to high-voltage power lines. He concluded that exposure to electromagnetic fields causes cancer.

(f) Many people are touting the benefits of the so-called "Mediterranean" diet because it apparently lowers the risk of heart disease. Residents of countries like Italy and Greece, for example, consume large amounts of vegetables and olive oil and suffer from heart problems at a much lower rate than Americans.

(g) My daughter came down with what appeared to be a run-of-the-mill case of the flu: fever, chills, congestion, sore throat. But it was a little weird. She was also experiencing really intense headaches and an extreme sensitivity to light. Those symptoms struck me as atypical of mere influenza, so I took her to the doctor. It's a good thing I did! It turns out she had a case of bacterial meningitis, which is so serious that it can cause brain damage if not treated early. Luckily, we caught it in time and she's doing fine.

# 8.

MATTHEW KNACHEL

## Inductive Logic II: Probability and Statistics

### I. The Probability Calculus

Inductive arguments, recall, are arguments whose premises support their conclusions insofar as they make them more *probable*. The more probable the conclusion in light of the premises, the stronger the argument; the less probable, the weaker. As we saw in the last chapter, it is often impossible to say with any precision exactly how probable the conclusion of a given inductive argument is in light of its premises; often, we can only make relative judgments, noting that one argument is stronger than another, because the conclusion is more probable, without being able to specify just how much more probable it is.

Sometimes, however, it is possible to specify precisely how probable the conclusion of an inductive argument is in light of its premises. To do that, we must learn something about how to calculate probabilities; we must learn the basics of the probability calculus. This is the branch of mathematics dealing with probability computations.<sup>1</sup> We will cover its most fundamental rules and learn to perform simple calculations. After that preliminary work, we use the tools provided by the probability calculus to think about how to make decisions in the face of uncertainty, and how to adjust our beliefs in the light of evidence. We will consider the question of what it means to be rational when engaging in these kinds of reasoning activities.

Finally, we will turn to an examination of inductive arguments involving statistics. Such arguments are of course pervasive in public discourse. Building on what we learned about probabilities, we will cover some of the most fundamental statistical concepts. This will allow us to understand various forms of statistical reasoning—from different methods of hypothesis testing to sampling techniques. In addition, even a rudimentary understanding of basic statistical concepts and reasoning methods will put us in a good position to recognize the myriad ways in which statistics are misunderstood, misused, and deployed with the intent to manipulate and deceive. As Mark Twain said, “There are three kinds of lies: lies, damned lies, and statistics.”<sup>2</sup> Advertisers, politicians, pundits—everybody in the persuasion business—trot out statistical claims to bolster their arguments, and more often than not they are either deliberately or mistakenly committing some sort of fallacy. We will end with a survey of these sorts of errors.

But first, we examine the probability calculus. Our study of how to compute probabilities will divide neatly into two sections, corresponding to the two basic types of probability calculations one can make. There are, on the one hand, probabilities of multiple events all occurring—or, equivalently, multiple propositions all being true; call these *conjunctive occurrences*. We will first learn how to calculate the probabilities of conjunctive occurrences—that this event *and* this other event *and* some other event *and so on* will occur. On the other hand, there are probabilities that at least one of a set of alternative events will occur—or, equivalently, that at least one of a set of propositions will be true; call these *disjunctive occurrences*. In the second half of our examination of the probability calculus we will learn how to calculate the probabilities of disjoint occurrences—that this event *or* this other event *or* some other event *or...* will occur.

1. Don't freak out about the word 'calculus'. We're not doing derivatives and integrals here; we're using that word in a generic sense, as in 'a system for performing calculations', or something like that. Also, don't get freaked out about 'mathematics'. This is really simple, fifth-grade stuff: adding and multiplying fractions and decimals.
2. Twain attributes the remark to British Prime Minister Benjamin Disraeli, though it's not really clear who said it first.

## Conjunctive Occurrences

Recall from our study of sentential logic that conjunctions are, roughly, ‘and’-sentences. We can think of calculating the probability of conjunctive occurrences as calculating the probability that a particular conjunction is true. If you roll two dice and want to know your chances of getting “snake eyes” (a pair of ones), you’re looking for the probability that you’ll get a one on the first die *and* a one on the second.

Such calculations can be simple or slightly more complex. What distinguishes the two cases is whether or not the events involved are *independent*. Events are independent when the occurrence of one has no effect on the probability that any of the others will occur. Consider the dice mentioned above. We considered two events: one on die #1, and one on die #2. Those events are independent. If I get a one on die #1, that doesn’t affect my chances of getting a one on the second die; there’s no mysterious interaction between the two dice, such that what happens with one can affect what happens with the other. They’re independent.<sup>3</sup> On the other hand, consider picking two cards from a standard deck (and keeping them after they’re drawn).<sup>4</sup> Here are two events: the first card is a heart, the second card is a heart. Those events are *not* independent. Getting a heart on the first draw affects your chances of getting a second heart (it makes the second heart less likely).

When events are independent, things are simple. We calculate the probability of their conjunctive occurrence by multiplying the probabilities of their individual occurrences. This is the Simple Product Rule:

$$P(a \bullet b \bullet c \bullet \dots) = P(a) \times P(b) \times P(c) \times \dots$$

This rule is abstract; it covers all cases of the conjunctive occurrence of independent events. ‘a’, ‘b’, and ‘c’ refer to events; the ellipses indicate that there may be any number of them. When we write ‘P’ followed by something in parentheses, that’s just the probability of the thing in parentheses coming to pass. On the left-hand side of the equation, we have a bunch of events with dots in between them. The dot means the same thing it did in SL: it’s short for *and*. So this equation just tells us that to compute the probability of a and b and c (and however many others there are) occurring, we just multiply together the individual probabilities of those events occurring on their own.

Go back to the dice above. We roll two dice. What’s the probability of getting a pair of ones? The events—one on die #1, one on die #2—are independent, so we can use the Simple Product Rule and just multiply together their individual probabilities.

What are those probabilities? We express probabilities as numbers between 0 and 1. An event with a probability of 0 definitely won’t happen (a proposition with a probability of 0 is certainly false); an event with a probability of 1 definitely will happen (a proposition with a probability of 1 is certainly true). Everything else is a number in between: closer to 1 is more probable; closer to 0, less. So, how probable is it for a rolled die to show a one? There are six possible outcomes when you roll a die; each one is equally likely. When that’s the case, the probability of the particular outcome is just 1 divided by the number of possibilities. The probability of rolling a one is  $\frac{1}{6}$ .

So, we calculate the probability of rolling “snake eyes” as follows:

$$\begin{aligned} P(\text{one on die \#1} \bullet \text{one on die \#2}) &= P(\text{one on die \#1}) \times P(\text{one on die \#2}) \\ &= \frac{1}{6} \times \frac{1}{6} \\ &= .0278 \end{aligned}$$

3. If you think otherwise, you’re committing what’s known as the Gambler’s Fallacy. It’s surprisingly common. Go to a casino and you’ll see people committing it. Head to the roulette wheel, for example, where people can bet on whether the ball lands in a red or a black space. After a run of say, five reds in a row, somebody will commit the fallacy: “Red is hot! I’m betting on it again.” This person believes that the results of the previous spins somehow affect the probability of the outcome of the next one. But they don’t. Notice that an equally compelling (and fallacious) case can be made for black: “Five reds in a row? Black is due. I’m betting on black.”
4. A standard deck has 52 playing cards, equally divided among four suits (hearts, diamonds, clubs, and spades) with 13 different cards in each suit: Ace (A), 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, Jack (J), Queen (Q), and King (K).

If you roll two dice a whole bunch of times, you'll get a pair of one a little less than 3% of the time.

We noted earlier that if you draw two cards from a deck, two possible outcomes—first card is a heart, second card is a heart—are not independent. So we couldn't calculate the probability of getting two spades using the Simple Product Rule. We could only do that if we made the two events independent—if we stipulated that after drawing the first card, you put it (randomly) back into the deck, so you're picking at random from a full deck of cards each time. In that case, you've got a  $\frac{1}{4}$  chance of picking a heart each time, so the probability of picking two in a row would be  $\frac{1}{4} \times \frac{1}{4}$ —and the probability of picking three in a row would be  $\frac{1}{4} \times \frac{1}{4} \times \frac{1}{4}$ , and so on.

Of course the more interesting question—and the more practical one, if you're a card player looking for an edge—is the original one: what's the probability of, say, drawing three hearts assuming, as is the case in all real-life card games, that you keep the cards as you draw them? As we noted, these events—heart on the first card, heart on the second card, heart on the third card—are not independent, because each time you succeed in drawing a heart, that affects your chances (negatively) of drawing another one. Let's think about this effect in the current case. The probability of drawing the first heart from a well-shuffled, complete deck is simple:  $\frac{1}{4}$ . It's the subsequent hearts that are complicated. How much of an effect does success at drawing that first heart have on the probability of drawing the second one? Well, if we've already drawn one heart, the deck from which we're attempting to draw the second is different from the original, full deck: specifically, it's short the one card already drawn—so there are only 51 total—and it's got fewer hearts now—12 instead of the original 13. 12 out of the remaining 51 cards are hearts, then. So the probability of drawing a second heart, assuming the first one has already been picked, is  $\frac{12}{51}$ . If we succeed in drawing the second heart, what are our chances at drawing a third? Again, in this case, the deck is different: we're now down to 50 total cards, only 11 of which are hearts. So the probability of getting the third heart is  $\frac{11}{50}$ .

It's these fractions— $\frac{1}{4}$ ,  $\frac{12}{51}$ , and  $\frac{11}{50}$ —that we must multiply together to determine the probability of drawing three straight hearts while keeping the cards. The result is (approximately) .013—a lower probability than that of picking 3 straight hearts when the cards are not kept, but replaced after each selection:  $\frac{1}{4} \times \frac{1}{4} \times \frac{1}{4} = .016$  (approximately). This is as it should be: it's harder to draw three straight hearts when the cards are kept, because each success diminishes the probability of drawing another heart. The events are not independent.

In general, when events are not independent, we have to make the same move that we made in the three-hearts case. Rather than considering the stand-alone probability of a second and third heart—as we could in the case where the events were independent—we had to consider the probability of those events *assuming* that other events had already occurred. We had to ask what the probability was of drawing a second heart, given that the first one had already been drawn; then we asked after the probability of drawing the third heart, given that the first two had been drawn.

We call such probabilities—the likelihood of an event occurring assuming that others have occurred—*conditional probabilities*. When events are not independent, the Simple Product Rule does not apply; instead, we must use the General Product Rule:

$$P(a \bullet b \bullet c \bullet \dots) = P(a) \times P(b | a) \times P(c | a \bullet b) \times \dots$$

The term ' $P(b | a)$ ' stands for the conditional probability of b occurring, provided a already has. The term ' $P(c | a \bullet b)$ ' stands for the conditional probability of c occurring, provided a and b already have. If there were a fourth event, d, we would have a this term on the right-hand side of the equation: ' $P(d | a \bullet b \bullet c)$ '. And so on.

Let's reinforce our understanding of how to compute the probabilities of conjunctive occurrences with a sample problem:

There is an urn filled with marbles of various colors. Specifically, it contains 20 red marbles, 30 blue marbles, and 50 white marbles. If we select 4 marbles from the urn at random, what's the probability that all four will be blue, (a) if we replace each marble after drawing it, and (b) if we keep each marble after drawing it?

Let's let 'B1' stand for the event of picking a blue marble on the first selection; and we'll let 'B2', 'B3', and 'B4' stand for the events of picking blue on the second, third, and fourth selections, respectively. We want the probability of all of these events occurring:

$$P(B1 \bullet B2 \bullet B3 \bullet B4) = ?$$

If we replace each marble after drawing it, then the events are independent: selecting blue on one drawing doesn't



affect our chances of selecting blue on any other; for each selection, the urn has the same composition of marbles. Since the events are independent in this case, we can use the Simple Product Rule to calculate the probability:

$$P(B1 \bullet B2 \bullet B3 \bullet B4) = P(B1) \times P(B2) \times P(B3) \times P(B4)$$

And since there are 100 total marbles in the urn, and 30 of them are blue, on each selection we have a  $\frac{30}{100}$  (= .3) probability of picking a blue marble.

$$P(B1 \bullet B2 \bullet B3 \bullet B4) = .3 \times .3 \times .3 \times .3 = .0081$$

If we don't replace the marbles after drawing them, then the events are not independent: each successful selection of a blue marble affects our chances (negatively) of drawing another blue marble. When events are not independent, we need to use the General Product Rule:

$$P(B1 \bullet B2 \bullet B3 \bullet B4) = P(B1) \times P(B2 | B1) \times P(B3 | B1 \bullet B2) \times P(B4 | B1 \bullet B2 \bullet B3)$$

On the first selection, we have the full urn, so  $P(B1) = \frac{30}{100}$ . But for the second term in our product, we have the conditional probability  $P(B2 | B1)$ ; we want to know the chances of selecting a second blue marble on the assumption that the first one has already been selected. In that situation, there are only 99 total marbles left, and 29 of them are blue. For the third term in our product, we have the conditional probability  $P(B3 | B1 \bullet B2)$ ; we want to know the chances of drawing a third blue marble on the assumption that the first and second ones have been selected. In that situation, there are only 98 total marbles left, and 28 of them are blue. And for the final term— $P(B4 | B1 \bullet B2 \bullet B3)$ —we want the probability of a fourth blue marble, assuming three have already been picked; there are 27 left out of a total of 97.

$$P(B1 \bullet B2 \bullet B3 \bullet B4) = \frac{30}{100} \times \frac{29}{99} \times \frac{28}{98} \times \frac{27}{97} = .007 \text{ (approximately)}$$

## Disjunctive Occurrences

Conjunctions are (roughly) 'and'-sentences. Disjunctions are (roughly) 'or'-sentences. So we can think of calculating the probability of disjunctive occurrences as calculating the probability that a particular disjunction is true. If, for example, you roll a die and you want to know the probability that it will come up with an odd number showing, you're looking for the probability that you'll roll a one or you'll roll a three or you'll roll a five.

As was the case with conjunctive occurrences, such calculations can be simple or slightly more complex. What distinguishes the two cases is whether or not the events involved are *mutually exclusive*. Events are mutually exclusive when at most one of them can occur—when the occurrence of one precludes the occurrence of any of the others. Consider the die mentioned above. We considered three events: it comes up showing one, it comes up showing three, and it comes up showing five. Those events are mutually exclusive; at most one of them can occur. If I roll a one, that means I can't roll a three or a five; if I roll a three, that means I can't roll a one or a five; and so on. (At most one of them can occur; notice, it's possible that none of them occur.) On the other hand, consider the dice example from earlier: rolling two dice, with the events under consideration rolling a one on die #1 and rolling a one on die #2. These events are *not* mutually exclusive. It's not the case that at most one of them could happen; they could both happen—we could roll snake eyes.

When events are mutually exclusive, things are simple. We calculate the probability of their disjunctive occurrence by adding the probabilities of their individual occurrences. This is the Simple Addition Rule:

$$P(a \vee b \vee c \vee \dots) = P(a) + P(b) + P(c) + \dots$$

This rule exactly parallels the Simple Product Rule from above. We replace that rule's dots with wedges, to reflect the fact that we're calculating the probability of *disjunctive* rather than *conjunctive* occurrences. And we replace the multiplication signs with additions signs on the righthand side of the equation to reflect the fact that in such cases we add rather than multiply the individual probabilities.

Go back to the die above. We roll it, and we want to know the probability of getting an odd number. There are three mutually exclusive events—rolling a one, rolling a three, and rolling a five—and we want their disjunctive probability;

that's  $P(\text{one} \vee \text{three} \vee \text{five})$ . Each individual event has a probability of  $1/6$ , so we calculate the disjunctive occurrence with the Simple Addition Rule thus:

$$\begin{aligned} P(\text{one} \vee \text{three} \vee \text{five}) &= P(\text{one}) + P(\text{three}) + P(\text{five}) \\ &= 1/6 + 1/6 + 1/6 = 3/6 = 1/2 \end{aligned}$$

This is a fine result, because it's the result we knew was coming. Think about it: we wanted to know the probability of rolling an odd number; half of the numbers are odd, and half are even; so the answer *better* be  $1/2$ . And it is.

Now, when events are not mutually exclusive, the Simple Addition Rule cannot be used; its results lead us astray. Consider a very simple example: flip a coin twice; what's the probability that you'll get heads at least once? That's a disjunctive occurrence: we're looking for the probability that you'll get heads on the first toss *or* heads on the second toss. But these two events—heads on toss #1, heads on toss #2—are not mutually exclusive. It's not the case that at most one can occur; you could get heads on both tosses. So in this case, the Simple Addition Rule will give us screwy results. The probability of tossing heads is  $1/2$ , so we get this:

$$\begin{aligned} P(\text{heads on \#1} \vee \text{heads on \#2}) &= P(\text{heads on \#1}) + P(\text{heads on \#2}) \\ &= 1/2 + 1/2 = 1 \text{ [WRONG!]} \end{aligned}$$

If we use the Simple Addition Rule in this case, we get the result that the probability of throwing heads at least once is 1; that is, it's absolutely certain to occur. Talk about screwy! We're not guaranteed to get heads at least once; we could toss tails twice in a row.

In cases such as this, where we want to calculate the probability of the disjunctive occurrence of events that are not mutually exclusive, we must do so indirectly, using the following universal truth:

$$P(\text{success}) = 1 - P(\text{failure})$$

This formula holds for any event or combination of events whatsoever. It says that the probability of any occurrence (singular, conjunctive, disjunctive, whatever) is equal to 1 minus the probability that it does not occur. 'Success' = it happens; 'failure' = it doesn't. Here's how we arrive at the formula. For any occurrence, there are two possibilities: either it will come to pass or it will not; success or failure. It's absolutely certain that at least one of these two will happen; that is,  $P(\text{success} \vee \text{failure}) = 1$ . Success and failure are (obviously) mutually exclusive outcomes (they can't both happen). So we can express  $P(\text{success} \vee \text{failure})$  using the Simple Addition Rule:  $P(\text{success} \vee \text{failure}) = P(\text{success}) + P(\text{failure})$ . And as we've already noted,  $P(\text{success} \vee \text{failure}) = 1$ , so  $P(\text{success}) + P(\text{failure}) = 1$ . Subtracting  $P(\text{failure})$  from each side of the equation gives us our universal formula:  $P(\text{success}) = 1 - P(\text{failure})$ .

Let's see how this formula works in practice. We'll go back to the case of flipping a coin twice. What's the probability of getting at least one head? Well, the probability of succeeding in getting at least one head is just 1 minus the probability of failing. What does failure look like in this case? No heads; two tails in a row. That is, tails on the first toss *and* tails on the second toss. See that 'and' in there? (I italicized it.) This was originally a *disjunctive*-occurrence calculation; now we've got a *conjunctive* occurrence calculation. We're looking for the probability of tails on the first toss *and* tails on the second toss:

$$P(\text{tails on toss \#1} \bullet \text{tails on toss \#2}) = ?$$

We know how to do problems like this. For conjunctive occurrences, we need first to ask whether the events are independent. In this case, they clearly are. Getting tails on the first toss doesn't affect my chances of getting tails on the second. That means we can use the Simple Product Rule:

$$\begin{aligned} P(\text{tails on toss \#1} \bullet \text{tails on toss \#2}) &= P(\text{tails on toss \#1}) \times P(\text{tails on toss \#2}) \\ &= 1/2 \times 1/2 = 1/4 \end{aligned}$$

Back to our universally true formula:  $P(\text{success}) = 1 - P(\text{failure})$ . The probability of failing to toss at least one head is  $1/4$ . The probability of succeeding in throwing at least one head, then, is just  $1 - 1/4 = 3/4$ .

So, generally speaking, when we're calculating the probability of disjunctive occurrences and the events are not

5. This makes good sense. If you throw a coin twice, there are four distinct ways things could go: (1) you throw heads twice; (2) you throw heads the first time, tails the second; (3) you throw tails the first time, heads the second; (4) you throw tails twice. In three out of those four scenarios (all but the last), you've thrown at least one head.

mutually exclusive, we need to do so indirectly, by calculating the probability of the failure of any of the disjunctive occurrences to come to pass and subtracting that from 1. This has the effect of turning a disjunctive occurrence calculation into a conjunctive occurrence calculation: the failure of a disjunction is a conjunction of failures. This is a familiar point from our study of SL in Chapter 4. Failure of a disjunction is a negated disjunction; negated disjunctions are equivalent to conjunctions of negations. This is one of DeMorgan's Theorems:

$$\sim (p \vee q) \equiv \sim p \bullet \sim q$$

Let's reinforce our understanding of how to compute probabilities with another sample problem. This problem will involve *both* conjunctive and disjunctive occurrences.

There is an urn filled with marbles of various colors. Specifically, it contains 20 red marbles, 30 blue marbles, and 50 white marbles. If we select 4 marbles from the urn at random, what's the probability that all four will be the same color, (a) if we replace each marble after drawing it, and (b) if we keep each marble after drawing it? Also, what's the probability that at least one of our four selections will be red, (c) if we replace each marble after drawing it, and (d) if we keep each marble after drawing it?

This problem splits into two: on the one hand, in (a) and (b), we're looking for the probability of drawing four marbles of the same color; on the other hand, in (c) and (d), we want the probability that at least one of the four will be red. We'll take these two questions up in turn.

First, the probability that all four will be the same color. We dealt with a narrower version of this question earlier when we calculated the probability that all four selections would be blue. But the present question is broader: we want to know the probability that they'll all be the same color, not just one color (like blue) in particular, but any of the three possibilities—red, white, or blue. There are three ways we could succeed in selecting four marbles of the same color: all four red, all four white, or all four blue. We want the probability that one of these will happen, and that's a disjunctive occurrence:

$$P(\text{all 4 red} \vee \text{all 4 white} \vee \text{all 4 blue}) = ?$$

When we are calculating the probability of disjunctive occurrences, our first step is to ask whether the events involved are mutually exclusive. In this case, they clearly are. At most, one of the three events—all four red, all four white, all four blue—will happen (and probably none of them will); we can't draw four marbles and have them all be red *and* all be white, for example. Since the events are mutually exclusive, we can use the Simple Addition Rule to calculate the probability of their disjunctive occurrence:

$$P(\text{all 4 red} \vee \text{all 4 white} \vee \text{all 4 blue}) = P(\text{all 4 red}) + P(\text{all 4 white}) + P(\text{all 4 blue})$$

So we need to calculate the probabilities for each individual color—that all will be red, all white, and all blue—and add those together. Again, this is the kind of calculation we did earlier, in our first practice problem, when we calculated the probability of all four marbles being blue. We just have to do the same for red and white. These are calculations of the probabilities of conjunctive occurrences:

$$P(R1 \bullet R2 \bullet R3 \bullet R4) = ?$$

$$P(W1 \bullet W2 \bullet W3 \bullet W4) = ?$$

(a) If we replace the marbles after drawing them, the events are independent, and so we can use the Simple Product Rule to do our calculations:

$$P(R1 \bullet R2 \bullet R3 \bullet R4) = P(R1) \times P(R2) \times P(R3) \times P(R4)$$

$$P(W1 \bullet W2 \bullet W3 \bullet W4) = P(W1) \times P(W2) \times P(W3) \times P(W4)$$

Since 20 of the 100 marbles are red, the probability of each of the individual red selections is .2; since 50 of the marbles are white, the probability for each white selection is .5.

$$P(R1 \bullet R2 \bullet R3 \bullet R4) = .2 \times .2 \times .2 \times .2 = .0016$$

$$P(W1 \bullet W2 \bullet W3 \bullet W4) = .5 \times .5 \times .5 \times .5 = .0625$$

In our earlier sample problem, we calculated the probability of picking four blue marbles: .0081. Putting these together, the probability of picking four marbles of the same color:

$$P(\text{all 4 red} \vee \text{all 4 white} \vee \text{all 4 blue}) = P(\text{all 4 red}) + P(\text{all 4 white}) + P(\text{all 4 blue})$$

$$= .0016 + .0625 + .0081$$

$$= .0722$$

(b) If we don't replace the marbles after each selection, the events are not independent, and so we must use the General Product Rule to do our calculations. The probability of selecting four red marbles is this:

$$P(R1 \bullet R2 \bullet R3 \bullet R4) = P(R1) \times P(R2 | R1) \times P(R3 | R1 \bullet R2) \times P(R4 | R1 \bullet R2 \bullet R3)$$

We start with 20 out of 100 red marbles, so  $P(R1) = \frac{20}{100}$ . On the second selection, we're assuming the first red marble has been drawn already, so there are only 19 red marbles left out of a total of 99;  $P(R2 | R1) = \frac{19}{99}$ . For the third selection, assuming that two red marbles have been drawn, we have  $P(R3 | R1 \bullet R2) = \frac{18}{98}$ . And on the fourth selection, we have  $P(R4 | R1 \bullet R2 \bullet R3) = \frac{17}{97}$ .

$$P(R1 \bullet R2 \bullet R3 \bullet R4) = \frac{20}{100} \times \frac{19}{99} \times \frac{18}{98} \times \frac{17}{97} = .0012 \text{ (approximately)}$$

The same considerations apply to our calculation of drawing four white marbles, except that we start with 50 of those on the first draw:

$$P(W1 \bullet W2 \bullet W3 \bullet W4) = \frac{50}{100} \times \frac{49}{99} \times \frac{48}{98} \times \frac{47}{97} = .0587 \text{ (approximately)}$$

In our earlier sample problem, we calculated the probability of picking four blue marbles as .007. Putting these together, the probability of picking four marbles of the same color:

$$\begin{aligned} P(\text{all 4 red} \vee \text{all 4 white} \vee \text{all 4 blue}) &= P(\text{all 4 red}) + P(\text{all 4 white}) + P(\text{all 4 blue}) \\ &= .0012 + .0587 + .007 \\ &= .0669 \text{ (approximately)} \end{aligned}$$

As we would expect, there's a slightly lower probability of selecting four marbles of the same color when we don't replace them after each selection.

We turn now to the second half of the problem, in which we are asked to calculate the probability that at least one of the four marbles selected will be red. The phrase 'at least one' is a clue: this is a disjunctive occurrence problem. We want to know the probability that the first marble will be red *or* the second will be red *or* the third *or* the fourth:

$$P(R1 \vee R2 \vee R3 \vee R4) = ?$$

When our task is to calculate the probability of disjunctive occurrences, the first step is to ask whether the events are mutually exclusive. In this case, they are not. It's not the case that at most one of our selections will be a red marble; we could pick two or three or even four (we calculated the probability of picking four just a minute ago). That means that we can't use the Simple Addition Rule to make this calculation. Instead, we must calculate the probability indirectly, relying on the fact that  $P(\text{success}) = 1 - P(\text{failure})$ . We must subtract the probability that we don't select any red marbles from 1:

$$P(R1 \vee R2 \vee R3 \vee R4) = 1 - P(\text{no red marbles})$$

As is always the case, the failure of a disjunctive occurrence is just a conjunction of individual failures. Not getting any red marbles is failing to get a red marble on the first draw *and* failing to get one on the second draw *and* failing on the third *and* on the fourth:

$$P(R1 \vee R2 \vee R3 \vee R4) = 1 - P(\sim R1 \bullet \sim R2 \bullet \sim R3 \bullet \sim R4)$$

In this formulation, ' $\sim R1$ ' stands for the eventuality of *not* drawing a red marble on the first selection, and the other terms for not getting red on the subsequent selections. Again, we're just borrowing symbols from SL.

Now we've got a conjunctive occurrence problem to solve, and so the question to ask is whether the events  $\sim R1$ ,  $\sim R2$ , and so on are independent or not. And the answer is that it depends on whether we replace the marbles after drawing them or not.

If we replace the marbles after each selection, then failure to pick red on one selection has no effect on the probability of failing to select red subsequently. It's the same urn— with 20 red marbles out of 100—for every pick. In that case, we can use the Simple Product Rule for our calculation:

$$P(R1 \vee R2 \vee R3 \vee R4) = 1 - [P(\sim R1) \times P(\sim R2) \times P(\sim R3) \times P(\sim R4)]$$

Since there are 20 red marbles, there are 80 non-red marbles, so the probability of picking a color other than red on any given selection is .8.

$$\begin{aligned} P(R1 \vee R2 \vee R3 \vee R4) &= 1 - (.8 \times .8 \times .8 \times .8) \\ &= 1 - .4096 \\ &= .5904 \end{aligned}$$

If we don't replace the marbles after each selection, then the events are not independent, and we must use the General Product Rule for our calculation. The quantity that we subtract from 1 will be this:

$$P(\sim R1) \times P(\sim R2 | \sim R1) \times P(\sim R3 | \sim R1 \bullet \sim R2) \times P(\sim R4 | \sim R1 \bullet \sim R2 \bullet \sim R3) = ?$$

On the first selection, our chances of picking a non-red marble are  $\frac{80}{100}$ . On the second selection, assuming we chose a non-red marble the first time, our chances are  $\frac{79}{99}$ . And on the third and fourth selections, the probabilities are  $\frac{78}{98}$  and  $\frac{77}{97}$ , respectively. Multiplying all these together, we get .4033 (approximately), and so our calculation of the probability of getting at least one red marble looks like this:

$$P(R1 \vee R2 \vee R3 \vee R4) = 1 - .4033 = .5967 \text{ (approximately)}$$

We have a slightly better chance at getting a red marble if we don't replace them, since each selection of a non-red marble makes the urn's composition a little more red-heavy.

## EXERCISES

1. Flip a coin 6 times; what's the probability of getting heads every time?
2. Go into a racquetball court and use duct tape to divide the floor into four quadrants of equal area. Throw three super-balls in random directions against the walls as hard as you can. What's the probability that all three balls come to rest in the same quadrant?
3. You're at your grandma's house for Christmas, and there's a bowl of holiday-themed M&Ms— red and green ones only. There are 500 candies in the bowl, with equal number of each color. Pick one, note its color, then eat it. Pick another, note its color, and eat it. Pick a third, note its color, and eat it. What's the probability that you ate three straight red M&Ms?
4. You and two of your friends enter a raffle. There is one prize: a complete set of Ultra Secret Rare Pokémon cards. There are 1000 total tickets sold; only one is the winner. You buy 20, and your friends each buy 10. What's the probability that one of you wins those Pokémon cards?
5. You're a 75% free-throw shooter. You get fouled attempting a 3-point shot, which means you get 3 free-throw attempts. What's the probability that you make at least one of them?
6. Roll two dice; what's the probability of rolling a seven? How about an eight?
7. In my county, 70% of people believe in Big Foot. Pick three people at random. What's the probability that at least one of them is a believer in Big Foot?
8. You see these two boxes here on the table? Each of them has jelly beans inside. We're going to play a little game, at the end of which you have to pick a random jelly bean and eat it. Here's the deal with the jelly beans. You may not be aware of this, but food scientists are able to create jelly beans with pretty much any flavor you want—and many you don't want. There is, in fact, such a thing as vomit-flavored jelly beans.<sup>6</sup> Anyway, in one of my two boxes, there are 100 total jelly beans, 8 of which are vomit-flavored (the rest are normal fruit flavors). In the other box, I have 50 jelly beans, 7 of which are vomit-flavored. Remember, this all ends with you choosing a random jelly bean and eating it. But you have a choice between two methods of determining how it will go down: (a) You flip a coin, and the result of the flip determines which of the two boxes you choose a jelly bean from; (b) I dump all the jelly beans into the same box and you pick from that. Which option do you choose? Which one minimizes the probability that you'll end up eating a vomit-flavored jelly bean? Or does it not make any difference?
9. For men entering college, the probability that they will finish a degree within four years is .329; for women, it's .438. Consider two freshmen—Albert and Betty. What's the probability that at least one of them will *fail* to complete college in at least four years? What's the probability that exactly one of them will *succeed* in doing so?
10. I love Chex Mix. My favorite things in the mix are those little pumpernickel chips. But they're relatively rare

6. Really: <http://mentalfloss.com/article/62593/how-doesjellybellycreateitsweirdflavors>

compared to the other ingredients. That's OK, though, since my second-favorite are the Chex pieces themselves, and they're pretty abundant. I don't know what the exact ratios are, but let's suppose that it's 50% Chex cereal, 30% pretzels, 10% crunchy bread sticks, and 10% my beloved pumpernickel chips. Suppose I've got a big bowl of Chex Mix: 1,000 total pieces of food. If I eat three pieces from the bowl, (a) what's the probability that at least one of them will be a pumpernickel chip? And (b) what's the probability that either all three will be pumpernickel chips or all three will be my second-favorite—Chex pieces?

11. You're playing draw poker. Here's how the game works: a poker hand is a combination of five cards; some combinations are better than others; in draw poker, you're dealt an initial hand, and then, after a round of wagering, you're given a chance to discard some of your cards (up to three) and draw new ones, hoping to improve your hand; after another round of betting, you see who wins. In this particular hand, you're initially dealt a 7 of hearts and the 4, 5, 6, and King of spades. This hand is quite weak on its own, but it's very close to being quite strong, in two ways: it's close to being a "flush", which is five cards of the same suit (you have four spades); it's also close to being a "straight", which is five cards of consecutive rank (you have four in a row in the 4, 5, 6, and 7). A flush beats a straight, but in this situation that doesn't matter; based on how the other players acted during the first round of betting, you're convinced that either the straight or the flush will win the money in the end. The question is, which one should you go for? Should you discard the King, hoping to draw a 3 or an 8 to complete your straight? Or should you discard the 7 of hearts, hoping to draw a spade to complete your flush? What's the probability for each? You should pick whichever one is higher.<sup>7</sup>

## II. Probability and Decision-Making: Value and Utility

The future is uncertain, but we have to make decisions every day that have an effect on our prospects, financial and otherwise. Faced with uncertainty, we do not merely throw up our hands and guess randomly about what to do; instead, we assess the potential risks and benefits of a variety of options, and choose to act in a way that maximizes the probability of a beneficial outcome. Things won't always turn out for the best, but we have to try to increase the chances that they will. To do so, we use our knowledge—or at least our best estimates—of the probabilities of future events to guide our decisions.

The process of decision-making in the face of uncertainty is most clearly illustrated with examples involving financial decisions. When we make a financial investment, or—what's effectively though not legally the same thing—a wager, we're putting money at risk with the hope that it will pay off in a larger sum of money in the future. We need a way of deciding whether such bets are good ones or not. Of course, we can evaluate our financial decisions in hindsight, and deem the winning bets good choices and the losing ones bad choices, but that's not a fair standard. The question is, knowing what we knew at the time we made our decision, did we make the choice that maximized the probability of a profitable outcome, even if the profit was not guaranteed?

To evaluate the soundness of a wager or investment, then, we need to look not at its worth after the fact—its final value, we might say—but rather at the value we can reasonably expect it to have in the future, based on what we know at the time the decision is made. We'll call this the *expected value*. To calculate the expected value of a wager or investment, we must take into consideration (a) the various possible ways in which the future might turn out that are relevant to our bet, (b) the value of our investment in those various circumstances, and (c) the probabilities that these various circumstances will come to pass. The expected value is a weighted average of the values in the different circumstances; it is weighted by the probabilities of each circumstance. Here is how we calculate expected value (EV):

$$EV = P(O_1) \times V(O_1) + P(O_2) \times V(O_2) + \dots + P(O_n) \times V(O_n)$$

This formula is a sum; each term in the sum is the product of a probability and a value. The terms ' $O_1$ ,  $O_2$ , ...,  $O_n$ ' refer to

7. Inspired by an exercise from Copi and Cohen, pp. 596 – 597

all the possible future outcomes that are relevant to our bet.  $P(O_x)$  is the probability that outcome #x will come to pass.  $V(O_x)$  is the value of our investment should outcome #x come to pass.

Perhaps the simplest possible scenario we can use to illustrate how this calculation works is the following: you and your roommate are bored, so you decide to play a game; you'll each put up a dollar, then flip a coin; if it comes up heads, you win all the money; if it comes up tails, she does.<sup>8</sup> What's the expected value of your \$1 bet? First, we need to consider which possible future circumstances are relevant to your bet's value. Clearly, there are two: the coin comes up heads, and the coin comes up tails. There are two outcomes in our formula:  $O_1$  = heads,  $O_2$  = tails. The probability of each of these is  $1/2$ . We must also consider the value of your investment in each of these circumstances. If heads comes up, the value is \$2—you keep your dollar and snag hers, too. If tails comes up, the value is \$0—you look on in horror as she pockets both bills. (Note: value is different from profit. You make a profit of \$1 if heads comes up, and you suffer a loss of \$1 if tails does—or your profit is -\$1. Value is how much money you're holding at the end.) Plugging the numbers into the formula, we get the expected value:

$$EV = P(\text{heads}) \times V(\text{heads}) + P(\text{tails}) \times V(\text{tails}) = \frac{1}{2} \times \$2 + \frac{1}{2} \times \$0 = \$1$$

The expected value of your \$1 bet is \$1. You invested a dollar with the expectation of a dollar in return. This is neither a good nor a bad bet. A good bet is one for which the expected value is greater than the amount invested; a bad bet is one for which it's less.

This suggests a standard for evaluating financial decisions in the real world: people should look to put their money to work in such a way that the expected value of their investments is as high as possible (and, of course, higher than the invested amount). Suppose I have \$1,000 free to invest. One way to put that money to work would be to stick it in a money market account, which is a special kind of savings deposit account one can open with a bank. Such accounts offer a return on your investment in the form of a payment of a certain amount of interest—a percentage of your deposit amount. Interest is typically specified as a yearly rate. So a money market account offering a 1% rate pays me 1% of my deposit amount after a year.<sup>9</sup> Let's calculate the expected value of an investment of my \$1,000 in such an account. We need to consider the possible outcomes that are relevant to my investment. I can only think of two possibilities: at the end of the year, the bank pays me my money; or, at the end of the year, I get stiffed—no money. The calculation looks like this:

$$EV = P(\text{paid}) \times V(\text{paid}) + P(\text{stiffed}) \times V(\text{stiffed})$$

One of the things that makes this kind of investment attractive is that it's virtually risk-free. Bank deposits of up to \$250,000 are insured by the federal government.<sup>10</sup> So even if the bank goes out of business before I withdraw my money, I'll still get paid in the end.<sup>11</sup> That means  $P(\text{paid}) = 1$  and  $P(\text{stiffed}) = 0$ . Nice. What's the value when I get paid? It's the initial \$1,000 plus the 1% interest. 1% of \$1,000 is \$10, so  $V(\text{paid}) = \$1,010$ .

That's not much of a return, but interest rates are low these days, and it's not a risky investment. We could increase the expected value if we put our money into something that's not a sure thing. One option is corporate bonds. For this type of investment, you lend your money to a company for a specified period of time (and they use it to build a factory or something), then you get paid back the principal investment plus some interest.<sup>12</sup> Corporate bonds are a riskier investment than bank deposits because they're not insured by the federal government. If your company goes bankrupt before the date you're supposed to get paid back, you lose your money. That is,  $P(\text{paid})$  in the expected value calculation

8. In this and what follows, I am indebted to Copi and Cohen's presentation for inspiration.

9. It's more complicated than this, but we're simplifying to make things easier.

10. They're insured through the FDIC—Federal Deposit Insurance Corporation—created during the Great Depression to prevent bank runs. Before this government insurance on deposits, if people thought a bank was in trouble, everybody tried to withdraw their money at the same time; that's a "bank run". Think about the scene in *It's a Wonderful Life* when George is about to leave on his honeymoon, but he has to go back to the Bailey Building and Loan to prevent such a catastrophe. Anyway, if everybody knows they'll get their money back even if the bank goes under, such things won't happen; that's what the FDIC is for.

11. Unless, of course, the federal government goes out of business. But in that case, \$1,000 might not be too useful anyway

above is no longer 1;  $P(\text{stiffed})$  is somewhere north of 0. What are the relevant probabilities? Well, it depends on the company. There are firms in the “credit rating” business—Moody’s, S&P, Fitch, etc.—that put out reports and classify companies according to how risky they are to loan money to. They assign ratings like ‘AAA’ (or ‘Aaa’, depending on the agency), ‘AA’, ‘BBB’, ‘CC’, and so on. The further into the alphabet you get, the higher the probability you’ll get stiffed. It’s impossible to say exactly what that probability is, of course; the credit rating agencies provide a rough guide, but ultimately it’s up to the individual investor to decide what the risks are and whether they’re worth it.

To determine whether the risks are worth it, we must compare the expected value of an investment in a corporate bond with a baseline, risk-free investment—like our money market account above. Since the probability of getting paid is less than 1, we must have a higher yield than 1% to justify choosing the corporate bond over the safer investment. How much higher? It depends on the company; it depends on how likely it is that we’ll get paid back in the end.

The expected value calculation is simple in these kinds of cases. Even though  $P(\text{stiffed})$  is not 0,  $V(\text{stiffed})$  is; if we get stiffed, our investment is worth nothing. So when calculating expected value, we can ignore the second term in the sum. All we have to do is multiply  $P(\text{paid})$  by  $V(\text{paid})$ .

Suppose we’re considering investing in a really reliable company; let’s say  $P(\text{paid}) = .99$ . Doing the math, in order for a corporate bond with this reliable company to be a better bet than a money market account, they’d have offer an interest rate of a little more than 2%. If we consider a less reliable company—say one for which  $P(\text{paid}) = .95$ —then we’d need a rate of little more than

when George is about to leave on his honeymoon, but he has to go back to the Bailey Building and Loan to prevent such a catastrophe. Anyway, if everybody knows they’ll get their money back even if the bank goes under, such things won’t happen; that’s what the FDIC is for.

Unless, of course, the federal government goes out of business. But in that case, \$1,000 is useful maybe as emergency toilet paper; I need canned goods and ammo at that point.

Again, there are all sorts of complications we’re glossing over to keep things simple.

Probably. There are different kinds of bankruptcies and lots of laws governing them; it’s possible for investors to get some money back in probate court. But it’s complicated. One thing’s for sure: our measly \$1,000 imaginary investment makes us too small-time to have much of a chance of getting paid during bankruptcy proceedings.<sup>14</sup> Historical data on the probability of default for companies at different ratings by agency are available.

6.3% to make this a better investment. If we go down to a 90% chance of getting paid back, we need a yield of more than 12% to justify that decision.

What does it mean to be a good, rational economic agent? How should a person, generally speaking, invest money? As we mentioned earlier, a plausible rule governing such decisions would be something like this: always choose the investment for which expected value is maximized.

But real people deviate from this rule in their monetary decisions, and it’s not at all clear that they’re irrational to do so. Consider the following choice: (a) we’ll flip a coin, and if it comes up heads, you win \$1,000, but if it comes up tails, you win nothing; (b) no coin flip, you just win \$499, guaranteed. The expected value of choice (b) is just the guaranteed \$499. The value of choice (a) can be easily calculated:

$$\begin{aligned} EV &= P(\text{heads}) \times V(\text{heads}) + P(\text{tails}) \times V(\text{tails}) \\ &= (.5 \times \$1,000) + (.5 \times \$0) \\ &= \$500 \end{aligned}$$

So according to our principle, it’s always rational to choose (a) over (b):  $\$500 > \$499$ . But in real life, most people who are offered such a choice go with the sure-thing, (b). (If you don’t share that intuition, try upping the stakes—coin flip for \$10,000 vs. \$4,999 for sure.) Are people who make such a choice behaving irrationally?

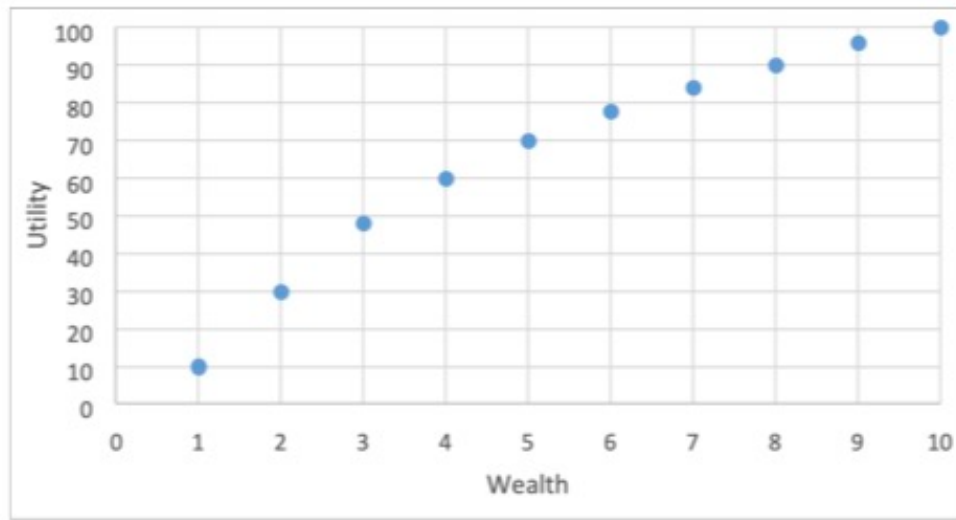
Not necessarily. What such examples show is that people take into consideration not merely the value, in dollars, of various choices, but the subjective significance of their outcomes—the degree to which they contribute to the person’s overall well-being. As opposed to ‘value’, we use the term ‘utility’ to refer to such considerations. In real life decisions, what matters is not the expected *value* of an investment choice, but its expected *utility*—the degree to which it satisfies a person’s desires, comports with subjective preferences.

The tendency of people to accept a sure thing over a risky wager, despite a lower expected value, is referred to as



*risk aversion*. This is the consequence of an idea first formalized by the mathematician Daniel Bernoulli in 1738: the diminishing marginal utility of wealth. The basic idea is that as the amount of money one has increases, each addition to one's fortune becomes less important, from a personal, subjective point of view. An extra \$1,000 means very little to Bill Gates; an extra \$1,000 for a poor college student would mean quite a lot. The money would add very little utility for Gates, but much more for the college student. Increases in one's fortune above zero mean more than subsequent increases. Bernoulli's utility function looked something like this<sup>12</sup>:

Diminishing Marginal Utility of Wealth



This explains the choice of the \$499 sure-thing over the coin flip for \$1,000. The utility attached to those first \$499 is greater than the extra utility of the additional possible \$501 dollars one could potentially win, so people opt to lock in the gain. Utility rises quickly at first, but levels out at higher amounts. From Bernoulli's chart, the utility of the sure-thing is somewhere around 70, while the utility of the full \$1,000 is only 30 more—100. Computing the expected utility of the coin-flip wager gives us this result:

$$\begin{aligned}
 EU &= P(\text{heads}) \times U(\text{heads}) + P(\text{tails}) \times U(\text{tails}) \\
 &= (.5 \times 100) + (.5 \times 0) \\
 &= 50
 \end{aligned}$$

The utility of 70 for the sure-thing easily beats the expected utility from the wager. It is possible to get people to accept risky bets over sure-things, but one must take into account this diminishing marginal utility. For a person whose personal utility function is like Bernoulli's, an offer of a mere \$300 (where the utility is down closer to 50) would make the decision more difficult. An offer of \$200 would cause them to choose the coin flip.

It has long been accepted economic doctrine that rational economic agents act in such a way as to maximize their utility, not their value. It is a matter of some dispute what sort of utility function best captures rational economic agency. Different economic theories assume different versions of ideal rationality for the agents in their models.

12. This function maps 1 unit of wealth to 10 units of utility (never mind what those units are). 2 units of wealth produces 30 units of utility, and so on: 3 – 48; 4 – 60; 5 – 70; 6 – 78; 7 – 84; 8 – 90; 9 – 96; 10 – 100. This mapping comes from Daniel Kahneman, 2011, *Thinking, Fast and Slow*, New York: Farrar, Strauss, and Giroux, p. 273.

Recently, this practice of assuming perfect utility-maximizing rationality of economic agents has been challenged. While it's true that the economic models generated under such assumptions can provide useful results, as a matter of fact, the behavior of real people (*homo sapiens* as opposed to "*homo economicus*"—the idealized economic man of the models) departs in predictable ways from the utility-maximizing ideal. Psychologists—especially Daniel Kahneman and Amos Tversky— have conducted a number of experiments that demonstrate pretty conclusively that people regularly behave in ways that, by the lights of economic theory, are irrational. For example, consider the following two scenarios (go slowly; think about your choices carefully):

You have \$1,000. Which would you choose?

Coin flip. Heads, you win another \$1,000; tails, you win nothing.

An additional \$500 for sure.

You have \$2,000. Which would you choose?

Coin flip. Heads you lose \$1,000; tails, you lose nothing.

Lose \$500 for sure.<sup>13</sup>

According to the Utility Theory of Bernoulli and contemporary economics, the rational agent would choose option (b) in each scenario. Though they start in different places, for each scenario option (a) is just a coin flip between \$1,000 and \$2,000, while (b) is \$1,500 for sure. Because of the diminishing marginal utility of wealth, (b) is the utility-maximizing choice each time. But as a matter of fact, most people choose option (b) only in the first scenario; they choose option (a) in the second. (If you don't share this intuition, try upping the stakes.) It turns out that most people dislike losing more than they like winning, so the prospect of a guaranteed loss in 2(b) is repugnant. Another example: would you accept a wager on a coin flip, where heads wins you \$1,500, but tails loses you \$1,000? Most people would not. (Again, if you're different, try upping the stakes.) And this despite the fact that *clearly* expected value and utility point to accepting the proposition.

Kahneman and Tversky's alternative to Utility Theory is called "Prospect Theory". It accounts for these and many other observed regularities in human economic behavior. For example, people's willingness to overpay for a very small chance at a very large gain (lottery tickets); also, their willingness to pay a premium to eliminate small risks (insurance); their willingness to take on risk to avoid large losses; and so on.<sup>14</sup>

It's debatable whether the observed deviations from idealized utility-maximizing behavior are rational or not. The question "What is an ideally rational economic agent?" is not one that we can answer easily. That's a question for philosophers to grapple with. The question that economists are grappling with is whether, and to what extent, they must incorporate these psychological regularities into their models. Real people are not the utility-maximizers the models say they are. Can we get more reliable economic predictions by taking their actual behavior into account? Behavioral economics is the branch of that discipline that answers this question in the affirmative. It is a rapidly developing field of research.

## EXERCISES

1. You buy a \$1 ticket in a raffle. There are 1,000 tickets sold. Tickets are selected out of one of those big round drums at random. There are 3 prizes: first prize is \$500; second prize is \$200; third prize is \$100. What's the expected value of your ticket?
2. On the eve of the 2016 U.S. presidential election, the poll-aggregating website 538.com predicted that Donald Trump had a 30% chance of winning. It's possible to wager on these sorts of things, believe it or not (with bookmakers or in "prediction markets"). On election night, right before 8:00pm EST, the "money line" odds on a

13. For this and many other examples, see Kahneman 2011. 18 Again, see Kahneman 2011 for details.

14. Again, see Kahneman 2011 for details.

Trump victory were +475. That means that a wager of \$100 on Trump would earn \$475 in profit, for a total final value of \$575. Assuming the 538.com crew had the probability of a Trump victory right, what was the expected value of a \$100 wager at 8:00pm at the odds listed?

3. You're offered three chances to roll a one with a fair die. You put up \$10 and your challenger puts up \$10. If you succeed in rolling one even once, you win all the money; if you fail, your challenger gets all the money. Should you accept the challenge? Why or why not?
4. You're considering placing a wager on a horse race. The horse you're considering is a longshot; the odds are 19 to 1. That means that for every dollar you wager, you'd win \$19 in profit (which means \$20 total in your pocket afterwards). How probable must it be that the horse will win for this to be a good wager (in the sense that the expected value is greater than the amount bet)?
5. I'm looking for a good deal in the junk bond market. These are highly risky corporate bonds; the risk is compensated for with higher yields. Suppose I find a company that I think has a 25% chance of going bankrupt before the bond matures. How high of a yield do I need to be offered to make this a good investment (again, in the sense that the expected value is greater than the price of the investment)?
6. For someone with a utility function like that described by Bernoulli (see above), what would their choice be if you offered them the following two options: (a) coin flip, with heads winning \$8,000 and tails winning \$2,000; (b) \$5,000 guaranteed? Explain why they would make that choice, in terms of expected utility. How would increasing the lower prize on the coin-flip option change things, if at all? Suppose we increased it to \$3,000. Or \$4,000. Explain your answers.

### III. Probability and Belief: Bayesian Reasoning

The great Scottish philosopher David Hume, in his *An Enquiry Concerning Human Understanding*, wrote, "In our reasonings concerning matter of fact, there are all imaginable degrees of assurance, from the highest certainty to the lowest species of moral evidence. A wise man, therefore, proportions his belief to the evidence." Hume is making a very important point about a kind of reasoning that we engage in every day: the adjustment of beliefs in light of evidence. We believe things with varying degrees of certainty, and as we make observations or learn new things that bear on those beliefs, we make adjustments to our beliefs, becoming more or less certain accordingly. Or, at least, that's what we *ought* to do. Hume's point is an important one because too often people do not adjust their beliefs when confronted with evidence—especially evidence *against* their cherished opinions. One needn't look far to see people behaving in this way: the persistence and ubiquity of the beliefs, for example, that vaccines cause autism, or that global warming is a myth, despite overwhelming evidence to the contrary, are a testament to the widespread failure of people to proportion their beliefs to the evidence, to a general lack of "wisdom", as Hume puts it.

Here we have a reasoning process—adjusting beliefs in light of evidence—which can be done well or badly. We need a way to distinguish good instances of this kind of reasoning from bad ones. We need a logic. As it happens, the tools for constructing such a logic are ready to hand: we can use the probability calculus to evaluate this kind of reasoning.

Our logic will be simple: it will be a formula providing an abstract model of perfectly rational belief-revision. The formula will tell us how to compute a conditional probability. It's named after the 18<sup>th</sup> century English reverend who first formulated it: Thomas Bayes. It is called "Bayes' Law" and reasoning according to its strictures is called "Bayesian reasoning".

At this point, you will naturally be asking yourself something like this: "What on Earth does a theorem about probability have to do with adjusting beliefs based on evidence?" Excellent question; I'm glad you asked. As Hume mentioned in the quote we started with, our beliefs come with varying degrees of certainty. Here, for example, are three things I believe: (a)  $1 + 1 = 2$ ; (b) the earth is approximately 93 million miles from the sun (on average); (c) I am related to Winston Churchill. I've listed them in descending order: I'm most confident in (a), least confident in (c). I'm more confident in (a) than (b), since I can figure out that  $1 + 1 = 2$  on my own, whereas I have to rely on the testimony of others for the Earth-

to-Sun distance. Still, that testimony gives me a much stronger belief than does the testimony that is the source of (c). My relation to Churchill is apparently through my maternal grandmother; the details are hazy. Still, she and everybody else in the family always said we were related to him, so I believe it.

“Fine,” you’re thinking, “but what does this have to do with probabilities?” Our degrees of belief in particular claims can vary between two extremes: complete doubt and absolute certainty. We could assign numbers to those states: complete doubt is 0; absolute certainty is 1. Probabilities also vary between 0 and 1! It’s natural to represent degrees of beliefs as probabilities. This is one of the philosophical interpretations of what probabilities really are.<sup>15</sup> It’s the so-called “subjective” interpretation, since degrees of belief are subjective states of mind; we call these “personal probabilities”. Think of rolling a die. The probability that it will come up showing a one is  $\frac{1}{6}$ . One way of understanding what that means is to say that, before the die was thrown, the degree to which you believed the proposition that the die will come up showing one—the amount of confidence you had in that claim—was  $\frac{1}{6}$ . You would’ve had more confidence in the claim that it would come up showing an odd number—a degree of belief of  $\frac{1}{2}$ .

We’re talking about the process of revising our beliefs when we’re confronted with evidence. In terms of probabilities, that means raising or lowering our personal probabilities as warranted by the evidence. Suppose, for example, that I was visiting my grandmother’s hometown and ran into a friend of hers from way back. In the course of the conversation, I mention how grandma was related to Churchill. “That’s funny,” says the friend, “your grandmother always told me she was related to Mussolini.” I’ve just received some evidence that bears on my belief that I’m related to Churchill. I never heard this Mussolini claim before. I’m starting to suspect that my grandmother had an odd eccentricity: she enjoyed telling people that she was related to famous leaders during World War II. (I wonder if she ever claimed to be related to Stalin. FDR? Let’s pray Hitler was never invoked. And Hirohito would strain credulity; my grandma was clearly not Japanese.) In response to this evidence, if I’m being rational, I would revise my belief that I’m related to Winston Churchill: I would lower my personal probability for that belief; I would believe it less strongly. If, on the other hand, my visit to my grandma’s hometown produced a different bit of evidence—let’s say a relative had done the relevant research and produced a family genealogy tracing the relation to Churchill—then I would revise my belief in the other direction, increasing my personal probability, believing it more strongly.

Since belief-revision in this sense just involves adjusting probabilities, our model for how it works is just a means of calculating the relevant probabilities. That’s why our logic can take the form of an equation. We want to know how strongly we should believe something, given some evidence about it. That’s a conditional probability. Let ‘H’ stand for a generic hypothesis—something we believe to some degree or other; let ‘E’ stand for some evidence we discover. What we want to know is how to calculate  $P(H | E)$ —the probability of H given E, how strongly we should believe H in light of the discovery of E.

Bayes’ Law tells us how to perform this calculation. Here’s one version of the equation:<sup>16</sup>

$$P(H | E) = \frac{P(H) \times P(E | H)}{P(E)}$$

This equation has some nice features. First of all, the presence of ‘P(H)’ in the numerator is intuitive. This is often referred to as the “prior probability” (or “prior” for short); it’s the degree to which the hypothesis was believed *prior* to the discovery of the evidence. It makes sense that this would be part of the calculation: how strongly I believe in something now ought to be (at least in part) a function of how strongly I used to believe it. Second, ‘P(E | H)’ is a

15. There’s a whole literature on this. See this article for an overview: Hájek, Alan, “Interpretations of Probability”, The Stanford Encyclopedia of Philosophy (Winter 2012 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/win2012/entries/probability-interpret/>>.

16. It’s easy to derive this theorem, starting with the general product rule. We know  $P(E \bullet H) = P(E) \times P(H | E)$ , no matter what ‘E’ and ‘H’ stand for. A little algebraic manipulation gives us  $P(H | E) = P(E \bullet H) / P(E)$ . It’s a truth of logic that the expression ‘E • H’ is equivalent to ‘H • E’, so we can replace ‘P(E • H)’ with ‘P(H • E)’ in the numerator. And again, by the general product rule,  $P(H \bullet E) = P(H) \times P(E | H)$ —our final numerator.

useful item to have in the calculation, since it's often a probability that can be known. Notice, this is the reverse of the conditional probability we're trying to calculate: it's the probability of the evidence, assuming that the hypothesis is true (it may not be, but we assume it is, as they say, "for the sake of argument"). Consider an example: as you may know, being sick in the morning can be a sign of pregnancy; if this were happening to you, the hypothesis you'd be entertaining would be that you're pregnant, and the evidence would be vomiting in the morning. The conditional probability you're interested in is  $P(\text{pregnant} \mid \text{vomiting})$ —that is, the probability that you're pregnant, given that you've been throwing up in the morning. Part of using Bayes' Law to make this calculation involves the reverse of that conditional probability:  $P(\text{vomiting} \mid \text{pregnant})$ —the probability that you'd be throwing up in the morning, assuming (for the sake of argument) that you are in fact pregnant. And that's something we can just look up; studies have been done. It turns out that about 60% of women experience have morning sickness (to the point of throwing up) during the first trimester of pregnancy. There are lots of facts like this available. Did you know that a craving for ice is a potential sign of anemia? Apparently it is: 44% of anemia patients have the desire to eat ice. Similar examples are not hard to find. It's worth noting, in addition, that sometimes the reverse probability in question— $P(E \mid H)$ —is 1. In the case of a prediction made by a scientific hypothesis, this is so. Isaac Newton's theory of universal gravitation, for example, predicts that objects dropped from the same height will take the same amount of time to reach the ground, regardless of their weights (provided that air resistance is not a factor). This prediction is just a mathematical result of the equation governing gravitational attraction. So if  $H$  is Newton's theory and  $E$  is a bowling ball and a feather taking the same amount of time to fall, then  $P(E \mid H) = 1$ ; if Newton's theory is true, then it's a mathematical certainty that the evidence will be observed.<sup>17</sup> So this version of Bayes' Law is attractive because of both probabilities in the numerator:  $P(H)$ , the prior probability, is natural, since the adjusted degree of belief ought to depend on the prior degree of belief; and  $P(E \mid H)$  is useful, since it's a probability that we can often know precisely. The formula is also nice in that it comports well with our intuitions about how belief-revision ought to work. It does this in three ways.

First, we know that implausible hypotheses are hard to get people to believe; as Carl Sagan once put it, "Extraordinary claims require extraordinary evidence." Putting this in terms of personal probabilities, an implausible hypothesis—and extraordinary claim—is just one with a low prior:  $P(H)$  is a small fraction. Consider an example. In the immediate aftermath of the 2358 U.S. presidential election, some people claim that the election was rigged (possibly by Mars) in favor of General Krang by way of a massive computer hacking scheme that manipulated the vote totals in key precincts.<sup>22</sup> You have very little confidence in this hypothesis— and give it an extremely low prior probability—for lots of reasons, but two in particular: (a) Voting machines in individual precincts are not networked together, so any hacking scheme would have to be carried out on a machine-by-machine basis across hundreds—if not thousands—of precincts, an operation of almost impossible complexity; (b) An organization with practically unlimited financial resources and the strongest possible motivation for uncovering such a scheme—namely, the anti-Krang campaign—looked at the data and concluded there was nothing fishy going on. But none of this stops wishful-thinking anti-Krang voters from digging for evidence that in fact the fix had been in for Krang. When people approach you with this kind of evidence—look at these suspiciously high turnout numbers from a handful of precincts in rural Wisconsin!— your degree of belief in the hypothesis—that the Martians had hacked the election—barely budes. This is proper; again, extraordinary claims require extraordinary evidence, and future you isn't seeing it. This intuitive fact about how belief-revision is supposed to work is borne out by the equation for Bayes' Law. Implausible hypotheses have a low prior— $P(H)$  is a small fraction. It's hard to increase our degree of belief in such propositions— $P(H \mid E)$  doesn't easily rise—simply because we're multiplying by a low fraction in the numerator when calculating the new probability.

The math mirrors the actual mechanics of belief-revision in two more ways. Here's a truism: the more strongly predictive piece of evidence is for a given hypothesis, the more it supports that hypothesis when we observe it. We saw above that women who are pregnant experience morning sickness about 60% of the time; also, patients suffering from anemia crave ice (for some reason) 44% of the time. In other words, throwing up in the morning is more strongly predictive of pregnancy than ice-craving is of anemia. Morning sickness would increase belief in the hypothesis of

17. Provided you set things up carefully. Check out this video: <https://www.youtube.com/watch?v=E43-CfukEgs>.

pregnancy more than ice-craving would increase belief in anemia. Again, this banal observation is borne out in the equation for Bayes' Law. When we're calculating how strongly we should believe in a hypothesis in light of evidence— $P(H | E)$ —we always multiply in the numerator by the reverse conditional probability— $P(E | H)$ —the probability that you'd observe the evidence, assuming the hypothesis is true. For pregnancy/sickness, this means multiplying by .6; for anemia/ice-craving, we multiply by .44. In the former case, we're multiplying by a higher number, so our degree of belief increases more.

A third intuitive fact about belief-revision that our equation correctly captures is this: surprising evidence provides strong confirmation of a hypothesis. Consider the example of Albert Einstein's general theory of relativity, which provided a new way of understanding gravity: the presence of massive objects in a particular region of space affects the *geometry of space itself*, causing it to be curved in that vicinity. Einstein's theory has a number of surprising consequences, one of which is that because space is warped around massive objects, light will not travel in a straight line in those places.<sup>18</sup> In this example, H is Einstein's general theory of relativity, and E is an observation of light following a curvy path. When Einstein first put forward his theory in 1915, it was met with incredulity by the scientific community, not least because of this astonishing prediction. Light bending? Crazy! And yet, four years later, Arthur Eddington, an English astronomer, devised and executed an experiment in which just such an effect was observed. He took pictures of stars in the night sky, then kept his camera trained on the same spot and took another picture during an eclipse of the sun (the only time the stars would also be visible during the day). The new picture showed the stars in slightly different positions, because during the eclipse, their light had to pass near the sun, whose mass caused their path to be deflected slightly, just as Einstein predicted. As soon as Eddington made his results public, newspapers around the world announced the confirmation of general relativity and Einstein became a star. As we said, surprising results provide strong confirmation; hardly anything could be more surprising than light bending. We can put this in terms of personal probabilities. Bending light was the evidence, so  $P(E)$  represents the degree of belief someone would have in the proposition that light will travel a curvy path. This was a very low number before Eddington's experiments. When we use it to calculate how strongly we should believe in general relativity given the evidence that light in fact bends— $P(H | E)$ —it's in the denominator of our equation. Dividing by a very small fraction means multiplying by its reciprocal, which is a very large number. This makes  $P(H | E)$  go up dramatically. Again, the math mirrors actual reasoning practice.

So, our initial formulation of Bayes' Law has a number of attractive features; it comports well with our intuitions about how belief-revision actually works. But it is not the version of Bayes' Law that we will settle on to make actual calculations. Instead, we will use a version that replaces the denominator— $P(E)$ —with something else. This is because that term is a bit tricky. It's the prior probability of the evidence. That's another subjective state—how strongly you believed the evidence would be observed prior to its actual observation, or something like that. Subjectivity isn't a bad thing in this context; we're trying to figure out how to adjust subjective states (degrees of belief), after all. But the more of it we can remove from the calculation, the more reliable our results. As we discussed, the subjective prior probability for the hypothesis in question— $P(H)$ —belongs in our equation: how strongly we believe in something now ought to be a function of how strongly we used to believe in it. The other item in the numerator— $P(E | H)$ —is most welcome, since it's something we can often just look up—an objective fact. But  $P(E)$  is problematic. It makes sense in the case of light bending and general relativity. But consider the example where I run into my grandma's old acquaintance and she tells me about her claims to be related to Mussolini. What was my prior for that? It's not clear there even was one; the possibility probably never even occurred to me. I'd like to get rid of the present denominator and replace it with the kinds of terms I like—those in the numerator.

I can do this rather easily. To see how, it will be helpful to consider the fact that when we're evaluating a hypothesis in light of some evidence, there are often alternative hypotheses that it's competing with. Suppose I've got a funny looking rash on my skin; this is the evidence. I want to know what's causing it. I may come up with a number of possible explanations. It's winter, so maybe it's just dry skin; that's one hypothesis. Call it 'H<sub>1</sub>'. Another possibility: we've just

18. Or, it is travelling a straight line, just through a space that is curved. Same thing.

started using a new laundry detergent at my house; maybe I'm having a reaction.  $H_2$  = detergent. Maybe it's more serious, though. I get on the Google and start searching.  $H_3$  = psoriasis (a kind of skin disease). Then my hypochondria gets out of control, and I get really scared:  $H_4$  = leprosy. That's all I can think of, but it may not be any of those:  $H_5$  = some other cause.

I've got five possible explanations for my rash—five hypotheses I might believe in to some degree in light of the evidence. Notice that the list is exhaustive: since I added  $H_5$  (something else), one of the five hypotheses will explain the rash. Since this is the case, we can say with certainty that I have a rash and it's caused by the cold, or I have a rash and it's caused by the detergent, or I have a rash and it's caused by psoriasis, or I have a rash and it's caused by leprosy, or I have a rash and it's caused by something else. Generally speaking, when a list of hypotheses is exhaustive of the possibilities, the following is a truth of logic:

$$E \equiv (E \cdot H_1) \vee (E \cdot H_2) \vee \dots \vee (E \cdot H_n)$$

For each of the conjunctions, it doesn't matter what order you put the conjuncts, so this true, too:

$$E \equiv (H_1 \cdot E) \vee (H_2 \cdot E) \vee \dots \vee (H_n \cdot E)$$

Remember, we're trying to replace  $P(E)$  in the denominator of our formula. Well, if  $E$  is equivalent to that long disjunction, then  $P(E)$  is equal to the probability of the disjunction:

$$P(E) = P[(H_1 \cdot E) \vee (H_2 \cdot E) \vee \dots \vee (H_n \cdot E)]$$

We're calculating a disjunctive probability. If we assume that the hypotheses are mutually exclusive (only one of them can be true), then we can use the Simple Addition Rule:<sup>19</sup>

$$P(E) = P(H_1 \cdot E) + P(H_2 \cdot E) + \dots + P(H_n \cdot E)$$

Each item in the sum is a conjunctive probability calculation, for which we can use the General Product Rule:

$$P(E) = P(H_1) \times P(E | H_1) + P(H_2) \times P(E | H_2) + \dots + P(H_n) \times P(E | H_n)$$

And look what we have there: each item in the sum is now a product of exactly the two types of terms that I like—a prior probability for a hypothesis, and the reverse conditional probability of the evidence assuming the hypothesis is true (the thing I can often just look up). I didn't like my old denominator, but it's equivalent to something I love. So I'll replace it. This is our final version of Bayes' Law:

	$P(H_k) \times P(E   H_k)$	
$P(H_k   E) =$	-----	$[1 \leq k \leq n]^{20}$
	$P(H_1) \times P(E   H_1) + P(H_2) \times P(E   H_2) + \dots + P(H_n) \times P(E   H_n)$	

Let's see how this works in practice. Consider the following scenario:

Your mom does the grocery shopping at your house. She goes to two stores: Fairsley Foods and Gibbons' Market. Gibbons' is closer to home, so she goes there more often—80% of the time. Fairsley sometimes has great deals, though, so she drives the extra distance and shops there 20% of the time.

You can't stand Fairsley. First of all, they've got these annoying commercials with the crazy owner shouting into the camera and acting like a fool. Second, you got lost in there once when you were a little kid and you've still got emotional scars. Finally, their produce section is terrible: in particular, their peaches—your favorite fruit—are often mealy and bland, practically inedible. In fact, you're so obsessed with good peaches that you made a study of it, collecting samples over a period of time from both stores, tasting and recording your data. It turns out that peaches from Fairsley are bad 40% of the time, while those from Gibbons' are only bad 20% of the time. (Peaches are a fickle fruit; you've got to expect some bad ones no matter how much care you take.)

19. I know. In the example, maybe it's the cold weather and the new detergent causing my rash. Let's set that possibility aside.

20. We add the subscript 'k' to the hypothesis we're entertaining, and stipulate the k is between 1 and n simply to ensure that the hypothesis in question is among the set of exhaustive, mutually exclusive possibilities  $H_1, H_2, \dots, H_n$ .

Anyway, one fine day you walk into the kitchen and notice a heaping mound of peaches in the fruit basket; mom apparently just went shopping. Licking your lips, you grab a peach and take a bite. Ugh! Mealy, bland—horrible. “Stupid Fairsley,” you mutter as you spit out the fruit. Question: is your belief that the peach came from Fairsley rational? How strongly should you believe that it came from that store?

This is the kind of question Bayes’ Law can help us answer. It’s asking us about how strongly we should believe in something; that’s just calculating a (conditional) probability. We want to know how strongly we should believe that the peach came from Fairsley; that’s our hypothesis. Let’s call it ‘F’. These types of calculations are always of *conditional* probabilities: we want the probability of the hypothesis *given* the evidence. In this case, the evidence is that the peach was bad; let’s call that ‘B’. So the probability we want to calculate is  $P(F | B)$ —the probability that the peach came from Fairsley given that it’s bad.

At this point, we reference Bayes’ Law and plug things into the formula. In the numerator, we want the prior probability for our hypothesis, and the reverse conditional probability of the evidence assuming the hypothesis is true:

$$P(F | B) = \frac{P(F) \times P(B | F)}{\text{-----}}$$

In the denominator, we need a sum, with each term in the sum having exactly the same form as our numerator: a prior probability for a hypothesis multiplied by the reverse conditional probability. The sum has to have one such term for each of our possible hypotheses. In our scenario, there are only two: that the fruit came from Fairsley, or that it came from Gibbons’. Let’s call the second hypothesis ‘G’. Our calculation looks like this:

$$P(F | B) = \frac{P(F) \times P(B | F)}{P(F) \times P(B | F) + P(G) \times P(B | G)}$$

Now we just have to find concrete numbers for these various probabilities in our little story. First,  $P(F)$  is the prior probability for the peach coming from Fairsley—that is, the probability that you would’ve assigned to it coming from Fairsley *prior* to discovering the evidence that it was bad—before you took a bite. Well, we know mom’s shopping habits: 80% of the time she goes to Gibbons’; 20% of the time she goes to Fairsley. So a random piece of food—our peach, for example—has a 20% probability of coming from Fairsley.  $P(F) = .2$ . And for that matter, the peach has an 80% probability of coming from Gibbons’, so the prior probability for that hypothesis— $P(G)$ —is .8. What about  $P(B | F)$ ? That’s the conditional probability that a peach will be bad assuming it came from Fairsley. We know that! You did a systematic study and concluded that 40% of Fairsley’s peaches are bad;  $P(B | F) = .4$ . Moreover, your study showed that 20% of peaches from Gibbons’ were bad, so  $P(B | G) = .2$ . We can now plug in the numbers and do the calculation:

$$P(F | B) = (.2 \times .4) / ((.2 \times .4) + (.8 \times .2)) = .08 / (.08 + .16) = 1/3$$

As a matter of fact, the probability that the bad peach you tasted came from Fairsley—the conclusion to which you jumped as soon as you took a bite—is only  $1/3$ . It’s twice as likely that the peach came from Gibbons’. Your belief is not rational. Despite the fact that Fairsley peaches are bad at twice the rate of Gibbons’, it’s far more likely that your peach came from Gibbons’, mainly because your mom does so much more of her shopping there.

So here we have an instance of Bayes’ Law performing the function of a logic—providing a method for distinguishing good from bad reasoning. Our little story, it turns out, depicted an instance of the latter, and Bayes’ Law showed that the reasoning was bad by providing a standard against which to measure it. Bayes’ Law, on this interpretation, is a model of perfectly rational belief-revision. Of course many real-life examples of that kind of reasoning can’t be subjected to the kind of rigorous analysis that the (made up) numbers in our scenario allowed. When we’re actually adjusting our beliefs in light of evidence, we often lack precise numbers; we don’t walk around with a calculator and an index card with Bayes’ Law on it, crunching the numbers every time we learn new things. Nevertheless, our actual practices ought to be informed by Bayesian principles; they ought to approximate the kind of rigorous process exemplified by the formula. We should keep in mind the need to be open to adjusting our prior convictions, the fact that alternative possibilities exist and ought to be taken into consideration, the significance of probability and uncertainty to our deliberations about what to believe and how strongly to believe it. Again, Hume: the wise person proportions belief according to the evidence.



## EXERCISES

1. Women are twice as likely to suffer from anxiety disorders as men: 8% to 4%. They're also more likely to attend college: these days, it's about a 60/40 ratio of women to men. (Are these two phenomena related? That's a question for another time.) If a random person is selected from my logic class, and that person suffers from an anxiety disorder, what's the probability that it's a woman?
2. Among Wisconsin residents, 80% are Green Bay Packers fans, 10% are Chicago Bears fans, and 10% favor some other football team (we're assuming every Wisconsinite has a favorite team). Packer fans aren't afraid to show their spirit: 75% of them wear clothes featuring the team logo. Bears fans are quite reluctant to reveal their loyalties in such hostile territory, so only 25% of them are obnoxious enough to wear Bears clothes. Fans of other teams aren't quite as scared: 50% of them wear their teams' gear. I've got a neighbor who does *not* wear clothes with his favorite team's logo. Suspicious (FIB?). What's the probability he's a Bears fan?
3. In my logic class, 20% of students are deadbeats: on exams, they just guess randomly. 60% of the students are pretty good, but unspectacular: they get correct answers 80% of the time. The remaining 20% of the students are geniuses: they get correct answers 100% of the time. I give a true/false exam. Afterwards, I pick one of the completed exams at random; the student got the first two questions correct. What's the probability that it's one of the deadbeats?

## IV. Basic Statistical Concepts and Techniques

In this section and the next, the goal is equip ourselves to understand, analyze, and criticize arguments using statistics. Such arguments are extremely common; they're also frequently manipulative and/or fallacious. As Mark Twain once said, "There are three kinds of lies: lies, damned lies, and statistics." It is possible, however, with a minimal understanding of some basic statistical concepts and techniques, along with an awareness of the various ways these are commonly misused (intentionally or not), to see the "lies" for what they are: bad arguments that shouldn't persuade us. In this section, we will provide a foundation of basic statistical knowledge. In the next, we will look at various statistical fallacies.

### *Averages: Mean vs. Median*

The word 'average' is slippery: it can be used to refer both to the arithmetic mean or the median of a set of values. The mean and median are often different, and when this is the case, use of the word 'average' is equivocal. A clever person can use this fact to her rhetorical advantage. We hear the word 'average' thrown around quite a bit in arguments: the average family has such-and-such an income, the average student carries such-and-such in student loan debt, and so on. Audiences are supposed to take this fictional average entity to be representative of all the others, and depending on the conclusion she's trying to convince people of, the person making the argument will choose between mean and median, picking the number that best serves her rhetorical purpose. It's important, therefore, for the critical listener to ask, every time the word 'average' is used, "Does this refer to the mean or the median? What's the difference between the two? How would using the other affect the argument?"

A simple example can make this clear.<sup>21</sup> I run a masonry contracting business on the side—Logical Constructions (a wholly owned subsidiary of LogiCorp). Including myself, 22 people work at Logical Constructions. This is how much

21. Inspiration for this example, as with much that follows, comes from Darrell Huff, 1954, *How to Lie with Statistics*, New York: Norton.

they're paid per year: \$350,000 for me (I'm the boss); \$75,000 each for two foremen; \$70,000 for my accountant; \$50,000 each for five stone masons; \$30,000 for the office secretary; \$25,000 each for two apprentices; and \$20,000 each for ten laborers. To calculate the mean salary at Logical Constructions, we add up all the individual salaries (my \$350,000, \$75,000 twice since there are two foremen, and so on) and divide by the number of employees. The result is \$50,000. To calculate the median salary, we put all the individual salaries in numerical order (ten entries of \$20,000 for the laborers, then two entries of \$25,000 for the apprentices, and so on) and find the middle number—or, as is the case with our set, which has an even number of entries, the mean of the middle two numbers. The middle two numbers are both \$25,000, so the median salary is \$25,000.

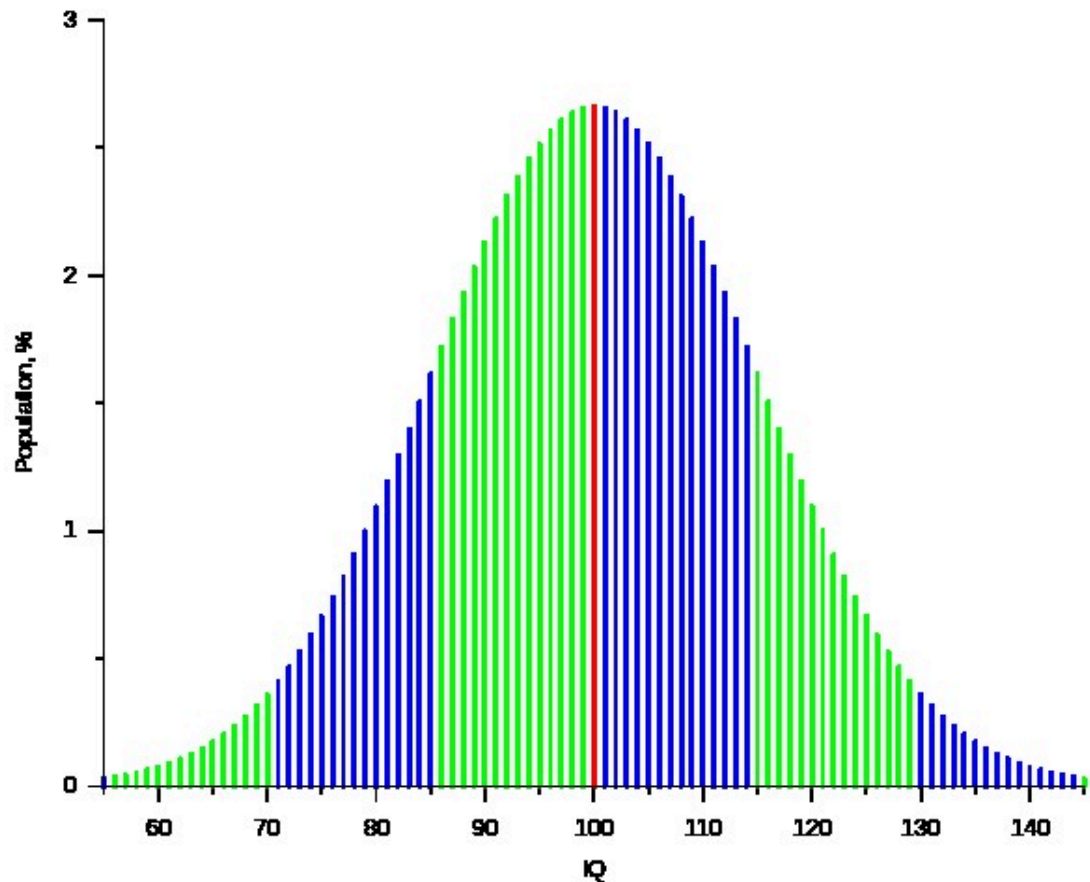
Now, you may have noticed, a lot of my workers don't get paid particularly well. In particular, those at the bottom—my ten laborers—are really getting the shaft: \$20,000 a year for that kind of back-breaking work is a raw deal. Suppose one day, as I'm driving past our construction site (in the back of my limo, naturally), I notice some outside agitators commiserating with my laborers during their (10-minute) lunch break—you know the type, union organizers, pinko commies (in this story, I'm a greedy capitalist; play along). They're trying to convince my employees to bargain collectively for higher wages. Now we have a debate: should the workers at Logical Constructions be paid more? I take one side of the issue; the workers and organizers take the other. In the course of making our arguments, we might both refer to the *average worker at Logical Constructions*. I'll want to do so in a way that makes it appear that this mythical worker is doing pretty well, and so we don't need to change anything; the organizers will want to do so in such a way that makes it appear that the average worker isn't do very well at all. We have two senses of 'average' to choose from: mean and median. In this case, the mean is higher, so I will use it: "The average worker at Logical Constructions makes \$50,000 per year. That's a pretty good wage!" My opponents, the union organizers, will counter, using the median: "The average worker at Logical Constructions makes a mere \$25,000 per year. Try raising a family on such a pittance!"

A lot hangs on which sense of 'average' we pick. This is true in lots of real-life circumstances. For example, household income in the United States is distributed much as salaries are at my fictional Logical Constructions company: those at the top of the range fare much better than those at the bottom.<sup>22</sup> In such circumstances, the mean is higher than the median. In 2014, the mean household income in the U.S. was \$72,641. That's pretty good! The median, however, was a mere \$53,657. That's a big difference! "The average family makes about \$72,000 per year" sounds a lot better than "The average family makes about \$53,000 per year."

### *Normal Distributions: Standard Deviation, Confidence Intervals*

If you gave IQ tests to a whole bunch of people, and then graphed the results on a histogram or bar chart—so that every time you saw a particular score, the bar for that score would get higher— you'd end up with a picture like this:

22. In 2014, the richest fifth of American households accounted for over 51% of income; the poorest fifth, 3%.



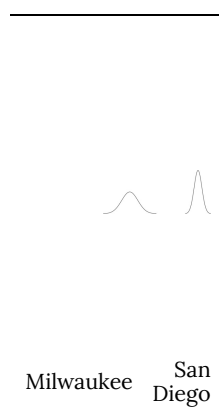
This kind of distribution is called a “normal” or “Gaussian” distribution<sup>23</sup>; because of its shape, it’s often called a “bell curve”. Besides IQ, many phenomena in nature are (approximately) distributed normally: height, blood pressure, motions of individual molecules in a collection, lifespans of industrial products, measurement errors, and so on.<sup>24</sup> And even when traits are not normally distributed, it can be useful to treat them as if they were. This is because the bell curve provides an extremely convenient starting point for making certain inferences. It’s convenient because one can know everything about such a curve by specifying two of its features: its mean (which, because the curve is symmetrical, is the same as its median) and its standard deviation.

We already understand the mean. Let’s get a grip on standard deviation. We don’t need to learn how to calculate it (though that can be done); we just want a qualitative (as opposed to quantitative) understanding of what it signifies. Roughly, it’s a measure of the *spread* of the data represented on the curve; it’s a way of indicating how far, on average, values tend to stray from the mean. An example can make this clear. Consider two cities: Milwaukee, WI and San Diego, CA. These two cities are different in a variety of ways, not least in the kind of weather their residents experience. Setting aside precipitation, let’s focus just on temperature. If you recorded the high temperatures every day in each town over

23. “Gaussian” because the great German mathematician Carl Friedrich Gauss made a study of such distributions in the early 19th century (in connection with their relationship to errors in measurement).

24. This is a consequence of a mathematical result, the Central Limit Theorem, the basic upshot of which is that if some random variable (a trait like IQ, for example, to be concrete) is the sum of many independent random variables (causes of IQ differences: lots of different genetic factors, lots of different environmental factors), then the variable (IQ) will be normally distributed. The mathematical theorem deals with abstract numbers, and the distribution is only perfectly “normal” when the number of independent variables approaches infinity. That’s why real-life distributions are only approximately normal.

a long period of time and made a histogram for each (with temperatures on the x-axis, number of days on the y-axis), you'd get two very different-looking curves. Maybe something like these:



The average high temperatures for the two cities—the peaks of the curves—would of course be different: San Diego is warmer on average than Milwaukee. But the *range* of temperatures experienced in Milwaukee is much greater than that in San Diego: some days in Milwaukee, the high temperature is below zero, while on some days in the summer it's over 100°F. San Diego, on the other hand, is basically always perfect: right around 70° or so.<sup>25</sup> The standard deviation of temperatures in Milwaukee is much greater than in San Diego. This is reflected in the shapes of the respective bell curves: Milwaukee's is shorter and wider—with a non-trivial number of days at the temperature extremes and a wide spread for all the other days—and San Diego's is taller and narrower—with temperatures hovering in a tight range all year, and hence more days at each temperature recorded (which explains the relative heights of the curves).

Once we know the mean and standard deviation of a normal distribution, we know everything we need to know about it. There are three very useful facts about these curves that can be stated in terms of the mean and standard deviation (SD). As a matter of mathematical fact, 68.3% of the population depicted on the curve (whether they're people with certain IQs, days on which certain temperatures were reached, measurements with a certain amount of error) falls within a range of one standard deviation on either side of the mean. So, for example, the mean IQ is 100; the standard deviation is 15. It follows that 68.3% of people have an IQ between 85 and 115—15 points (one SD) on either side of 100 (the mean). Another fact: 95.4% of the population depicted on a bell curve will fall within a range two standard deviations from the mean. So 95.4% of people have an IQ between 70 and 130—30 points (2 SDs) on either side of 100. Finally, 99.7% of the population falls within three standard deviations of the mean; 99.7% of people have IQs between 55 and 145. These ranges are called *confidence intervals*.<sup>26</sup> They are convenient reference points commonly used in statistical inference.<sup>27</sup>

25. This is an exaggeration, of course, but not much of one. The average high in San Diego in January is 65°; in July, it's 75°.

Meanwhile, in Milwaukee, the average high in January is 29°, while in July it's 80°.

26. Pick a person at random. How confident are you that they have an IQ between 70 and 130? 95.4%, that's how confident.

27. As a matter of fact, in current practice, other confidence intervals are more often used: 90%, (exactly) 95%, 99%, etc.

These ranges lie on either side of the mean within non-whole-number multiples of the standard deviation. For example, the exactly-95% interval is 1.96 SDs to either side of the mean. The convenience of calculators and spreadsheets to do our math for us makes these confidence intervals more practical. But we'll stick with the 68.3/95.4/99.7 intervals for simplicity's sake.

## *Statistical Inference: Hypothesis Testing*

If we start with knowledge of the properties of a given normal distribution, we can test claims about the world to which that information is relevant. Starting with a bell curve—information of a general nature—we can make draw conclusions about particular hypotheses. These are conclusions of inductive arguments; they are not certain, but more or less probable. When we use knowledge of normal distributions to draw them, we can be precise about how probable they are. This is inductive logic.

The basic pattern of the kinds of inferences we're talking about is this: one formulates a hypothesis, then runs an experiment to test it; the test involves comparing the results of that experiment to what is known (some normal distribution); depending on how well the results of the experiment comport with what would be expected given the background knowledge represented by the bell curve, we draw a conclusion about whether or not the hypothesis is true.

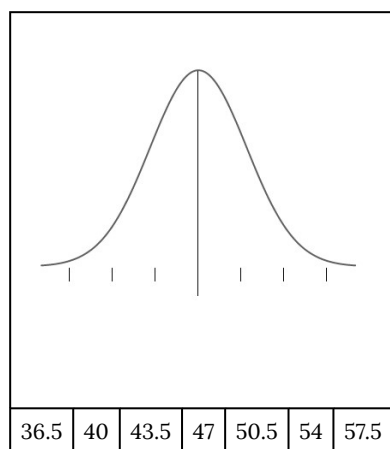
Though they are applicable in a very wide range of contexts, it's perhaps easiest to explain the patterns of reasoning we're going to examine using examples from medicine. These kinds of cases are vivid; they aid in understanding by making the consequences of potential errors more real. Also, in these cases the hypotheses being tested are relatively simple: claims about individuals' health—whether they're healthy or sick, whether they have some condition or don't—as opposed to hypotheses dealing with larger populations and measurements of their properties. Examining these simpler cases will allow us to see more clearly the underlying patterns of reasoning that cover all such instances of hypothesis testing, and to gain familiarity with the vocabulary statisticians use in their work.

The knowledge we start with is how some trait relevant to the particular condition is distributed in the population generally—a bell curve.<sup>28</sup> The experiment we run is to measure the relevant trait in the individual whose health we're assessing. The result of a comparison with the result of this measurement and the known distribution of the trait tells us something about whether or not the person is healthy. Suppose we start with information about how a trait is distributed among people who are healthy. Hematocrit, for example, is a measure of how much of a person's blood is taken up by red blood cells—expressed as a percentage (of total blood volume). Lower hematocrit levels are associated with anemia; higher levels are associated with dehydration, certain kinds of tumors, and other disorders. Among healthy men, the mean hematocrit level is 47%, with a standard deviation of 3.5%. We can draw the curve, noting the boundaries of the confidence intervals:

spreadsheets to do our math for us makes these confidence intervals more practical. But we'll stick with the 68.3/95.4/99.7 intervals for simplicity's sake.<sup>34</sup> Again, the actual distribution may not be normal, but we will assume that it is in our examples. The basic patterns of reasoning are similar when dealing with different kinds of distributions.

Hematocrit Levels, Healthy Men

28. Again, the actual distribution may not be normal, but we will assume that it is in our examples. The basic patterns of reasoning are similar when dealing with different kinds of distributions.



Because of the fixed mathematical properties of the bell curve, we know that 68.3% of healthy men have hematocrit levels between 43.5% and 50.5%; 95.4% of them are between 40% and 54%; and 99.7% of them are between 36.5% and 57.5%. Let's consider a man whose health we're interested in evaluating. Call him Larry. We take a sample of Larry's blood and measure the hematocrit level. We compare it to the values on the curve to see if there might be some reason to be concerned about Larry's health. Remember, the curve tells us the levels of hematocrit for *healthy* men; we want to know if Larry's one of them. The hypothesis we're testing is that Larry's healthy. Statisticians often refer the hypothesis under examination in such tests as the "null hypothesis"—a default assumption, something we're inclined to believe unless we discover evidence against it. Anyway, we're measuring Larry's hematocrit; what kind of result should he be hoping for? Clearly, he'd like to be as close to the middle, fat part of the curve as possible; that's where most of the healthy people are. The further away from the average healthy person's level of hematocrit he strays, the more he's worried about his health. That's how these tests work: if the result of the experiment (measuring Larry's hematocrit) is sufficiently close to the mean, we have no reason to reject the null hypothesis (that Larry's healthy); if the result is far away, we do have reason to reject it.

How far away from the mean is too far away? It depends. A typical cutoff is two standard deviations from the mean—the 95.4% confidence interval.<sup>29</sup> That is, if Larry's hematocrit level is below 40% or above 54%, then we might say we have reason to doubt the null hypothesis that Larry is healthy. The language statisticians use for such a result—say, for example, if Larry's hematocrit came in at 38%—is to say that it's "statistically significant". In addition, they specify the *level* at which it's significant—an indication of the confidence-interval cutoff that was used. In this case, we'd say Larry's result of 38% is *statistically significant at the .05 level*. ( $95\% = .95$ ;  $1 - .95 = .05$ ) Either Larry is unhealthy (anemia, most likely), or he's among the (approximately) 5% of healthy people who fall outside of the two standard-deviation range. If he came in at a level even further from the mean—say, 36%—we would say that this result is significant at the .003 level ( $99.7\% = .997$ ;  $1 - .997 = .003$ ). That would give us all the more reason to doubt that Larry is healthy.

So, when we're designing a medical test like this, the crucial decision to make is where to set the cutoff. Again, typically that's the 95% confidence interval. If a result falls outside that range, the person tests "positive" for whatever condition we're on the lookout for. (Of course, a "positive" result is hardly positive news—in the sense of being something you want to hear.) But these sorts of results are not conclusive: it may be that the null hypothesis (this person is healthy) is true, and that they're simply one of the relative rare 5% who fall on the outskirts of the curve. In such a case, we would say that the test has given the person a "false positive" result: the test indicates sickness when in fact there is none. Statisticians refer to this kind of mistake as "type I error". We could reduce the number of mistaken results our test gives by changing the confidence levels at which we give a positive result. Returning to the concrete example above: suppose Larry has a hematocrit level of 38%, but that he is not in fact anemic; since 38% is outside of the two

29. Actually, the typical level is now exactly 95%, or 1.96 standard deviations from the mean. From now on, we're just going to pretend that the 95.4% and 95% levels are the same thing.

standard-deviation range, our test would give Larry a false positive result if we used the 95% confidence level. However, if we raised the threshold of statistical significance to the three standard-deviation level of 99.7%, Larry would not get flagged for anemia; there would be no false positive, no type I error.

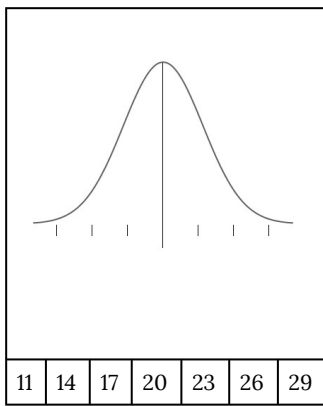
So we should always use the wider range on these kinds of tests to avoid false positives, right? Not so fast. There's another kind of mistake we can make: false negatives, or type II errors. Increasing our range increases our risk of this second kind of foul-up. Down there at the skinny end of the curve there are relatively few healthy people. Sick people are the ones who generally have measurements in that range; they're the ones we're trying to catch. When we issue a false negative, we're missing them. A false negative occurs when the test tells you there's no reason to doubt the null hypothesis (that you're healthy), when as a matter of fact you are sick. If we increase our range from two to three standard deviations—from the 95% level to the 99.7% level—we will avoid giving a false positive result to Larry, who is healthy despite his low 38% hematocrit level. But we will end up giving false reassurance to some anemic people who have levels similar to Larry's; someone who has a level of 38% and is sick will get a false negative result if we only flag those outside the 99.7% confidence interval (36.5% – 57.5%).

This is a perennial dilemma in medical screening: how best to strike a balance between the two types of errors—between needlessly alarming healthy people with false positive results and failing to detect sickness in people with false negative results. The terms clinicians use to characterize how well diagnostic tests perform along these two dimensions are *sensitivity* and *specificity*. A highly sensitive test will catch a large number of cases of sickness—it has a high rate of true positive results; of course, this comes at the cost of increasing the number of false positive results as well. A test with a high level of specificity will have a high rate of true negative results—correctly identifying healthy people as such; the cost of increased specificity, though, is an increase in the number of false negative results—sick people that the test misses. Since every false positive is a missed opportunity for a true negative, increasing sensitivity comes at the cost of decreasing specificity. And since every false negative is a missed true positive, increasing specificity comes at the cost of decreasing sensitivity. A final bit of medical jargon: a screening test is *accurate* to the degree that it is *both* sensitive and specific.

Given sufficiently thorough information about the distributions of traits among healthy and sick populations, clinicians can rig their diagnostic tests to be as sensitive or specific as they like. But since those two properties pull in opposite directions, there are limits to degree of accuracy that is possible. And depending on the particular case, it may be desirable to sacrifice specificity for more sensitivity, or *vice versa*.

To see how a screening test might be rigged to maximize sensitivity, let's consider an abstract hypothetical example. Suppose we knew the distribution of a certain trait among the population of people suffering from a certain disease. (Contrast this with our starting point above: knowledge of the distribution among *healthy* individuals.) This kind of knowledge is common in medical contexts: various so-called biomarkers—gene mutations, proteins in the blood, etc.—are known to be indicative of certain conditions; often, one can know how such markers are distributed among people with the condition. Again, keeping it abstract and hypothetical, suppose we know that among people who suffer from Disease X, the mean level of a certain biomarker  $\beta$  for the disease is 20, with a standard deviation of 3. We can sum up this knowledge with a curve:

$\beta$  Levels, People with Disease X



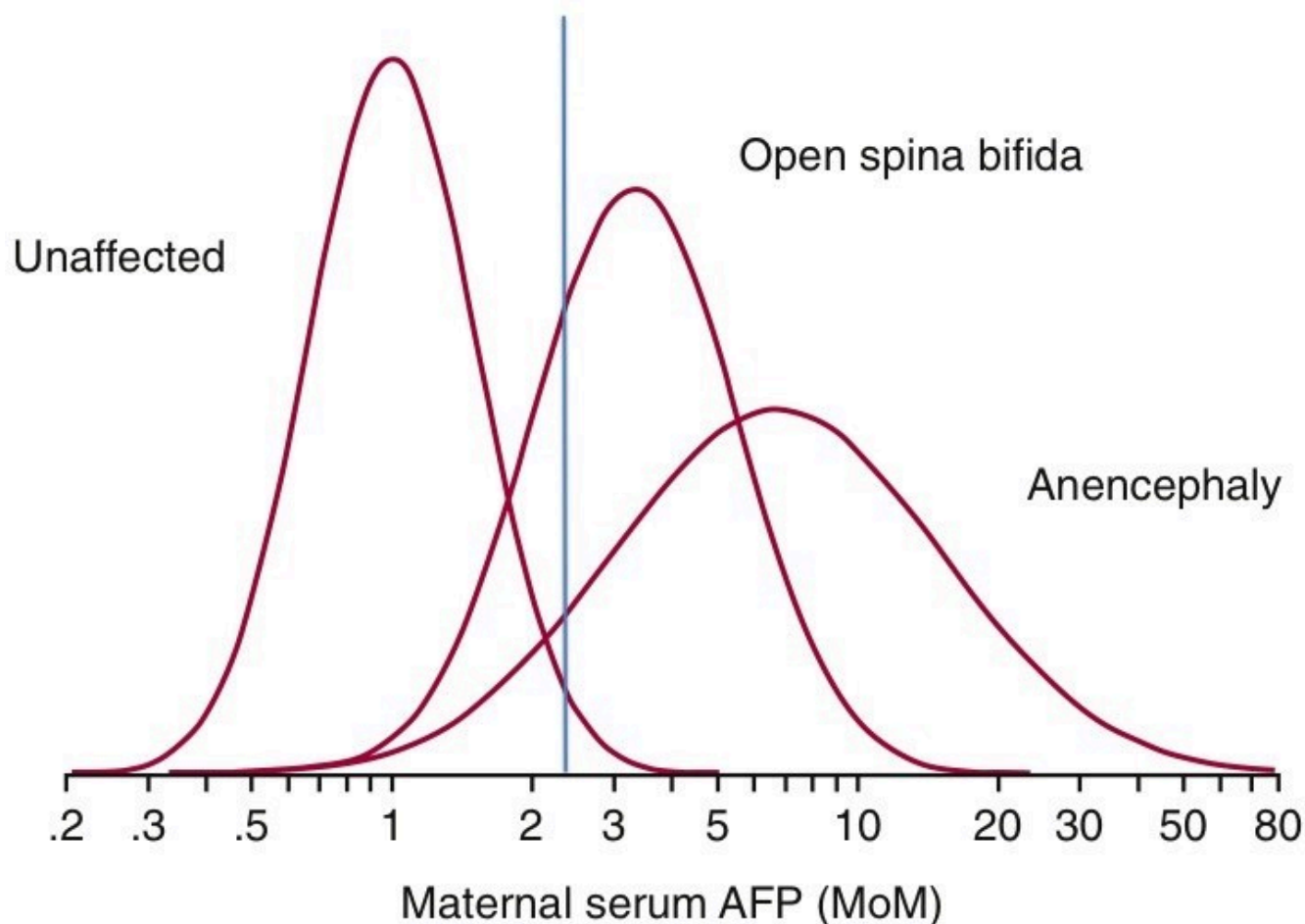
Now, suppose Disease X is very serious indeed. It would be a benefit to public health if we were able to devise a screening test that could catch as many cases as possible—a test with a high sensitivity. Given the knowledge we have about the distribution of  $\beta$  among patients with the disease, we can make our test as sensitive as we like. We know, as a matter of mathematical fact, that 68.3% percent of people with the disease have  $\beta$ -levels between 17 and 23; 95.4% of people with the disease have levels between 14 and 26; 99.7% have levels between 11 and 29. Given these facts, we can devise a test that will catch 99.7% of cases of Disease X like so: measure the level of biomarker  $\beta$  in people, and if they have a value between 11 and 29, they get a positive test result; a positive result is indicative of disease. This will catch 99.7% of cases of the condition, because the range chosen is three standard deviations on either side of the mean, and that range contains

99.7% of unhealthy people; if we flag everybody in that range, we will catch 99.7% of cases. Of course, we'll probably end up catching a whole lot of healthy people as well if we cast our net this wide; we'll get a lot of false positives. We could correct for this by making our test less sensitive, say by lowering the threshold for a positive test to the two standard-deviation range of 14 – 26. We would now only catch 95.4% of cases of sickness, but we would reduce the number of healthy people given false positives; instead, they would get true *negative* results, increasing the specificity of our test. Notice that the way we used the bell curve in our hypothetical test for Disease X was different from the way we used the bell curve in our test of hematocrit levels above. In that case, we flagged people as potentially sick when they fell *outside* of a range around the mean; in the new case, we flagged people as potentially sick when they fell *inside* a certain range. This difference corresponds to the differences in the two populations the respective distributions represent: in the case of hematocrit, we started with a curve depicting the distribution of a trait among healthy people; in the second case, we started with a curve telling us about sick people. In the former case, sick people will tend to be far from the mean; in the latter, they'll tend to cluster closer.

The tension we've noted between sensitivity and specificity—between increasing the number of cases our diagnostic test catches and reducing the number of false positives it produces, can be seen when show curves for healthy populations and sick populations in the same graph. There is a biomarker called alpha-fetoprotein in the blood serum of pregnant women. Low levels of this protein are associated with Down syndrome in the fetus; high levels are associated with neural tube defects like open spina bifida (spine isn't completely inside the body) and anencephaly (hardly any of the brain/skull develops). These are serious conditions—especially those associated with the high levels: if the baby has open spina bifida, you need to be ready for that (with specialists and special equipment) at the time of birth; in cases of anencephaly, the fetus will not be viable (at worst) or will live without sensation or awareness (at best?). Early in pregnancy, these conditions are screened for. Since they're so serious, you'd like to catch as many cases as possible. And yet, you'd like to avoid alarming false positive results for these conditions. The following chart, with bell curves for



healthy babies, those with open spina bifida, and anencephaly, illustrates the difficult tradeoffs in making these sorts of decisions<sup>30</sup>:



The vertical line at 2.5 MoM (multiples of the median) is the typical cutoff for a “positive” result (flagged for potential problems). On the one hand, there are substantial portions of the two curves representing the unhealthy populations—to the left of that line—that won’t be flagged by the test. Those are cases of sickness that we won’t catch—false negatives. On the other hand, there are a *whole lot* of healthy babies whose parents are going to be unnecessarily alarmed. The area of the “Unaffected” curve to the right of the line may not look like much, but these curves aren’t drawn on a linear scale. If they were, that curve would be *much (much!)* higher than the two for open spina bifida and anencephaly: those conditions are really rare; there are *far* more healthy babies. The upshot is, that tiny-looking portion of the healthy curve represents a lot of false positives.

Again, this kind of tradeoff between sensitivity and specificity often presents clinicians with difficult choices in designing diagnostic tests. They must weigh the benefits of catching as many cases as possible against the potential costs of too many false positives. Among the costs are the psychological impacts of getting a false positive. As a parent who experienced it, I can tell you getting news of potential open spina bifida or anencephaly is quite traumatic.<sup>31</sup> But it could be worse. For example, when a biomarker for AIDS was first identified in the mid-1980s, people at the Centers for Disease Control considered screening for the disease among the entire population. The test was sensitive,

30. Picture from a post at [www.pregnancylab.net](http://www.pregnancylab.net) by David Grenache, PhD: <http://www.pregnancylab.net/2012/11/screening-forneuraltubedefects.html>

31. False positive: the baby was perfectly healthy.

so they knew they would catch a lot of cases. But they also knew that there would be a good number of false positives. Considering the hysteria that would likely arise from so many diagnoses of the dreaded illness (in those days, people knew hardly anything about AIDS; people were dying of a mysterious illness, and fear and misinformation were widespread), they decided against universal screening. Sometimes the negative consequences of false positives include financial and medical costs. In 2015, the American Cancer Society changed its recommendations for breast-cancer screening: instead of starting yearly mammograms at age 40, women should wait until age 45.<sup>32</sup> This was a controversial decision. Afterwards, many women came forward to testify that their lives were saved by early detection of breast cancer, and that under the new guidelines they may not have fared so well. But against the benefit of catching those cases, the ACS had to weigh the costs of false-positive mammograms. The follow-up to a positive mammogram is often a biopsy; that's an invasive surgical procedure, and costly. Contrast that with the follow-up to a positive result for open spina bifida/anencephaly: a non-invasive, cheap ultrasound. And unlike an ultrasound, the biopsy is sometimes quite difficult to interpret; you get some diagnoses of cancer when cancer is not present. Those women may go on to receive treatment—chemotherapy, radiation—for cancer that they don't have. The costs and physical side effects of that are severe.<sup>33</sup> In one study, it was determined that for every life saved by mammography screening, there were 100 women who got false positives (and learned about it after a biopsy) and five women treated for cancer they didn't have.<sup>34</sup> The logic of statistical hypothesis testing is relatively clear. What's not clear is how we ought to apply those relatively straightforward techniques in actual practice. That often involves difficult financial, medical, and moral decisions.

### *Statistical Inference: Sampling*

When we were testing hypotheses, our starting point was knowledge about how traits were distributed among a large population—e.g., hematocrit levels among healthy men. We now ask a pressing question: how do we acquire such knowledge? How do we figure out how things stand with a very large population? The difficulty is that it's usually impossible to check every member of the population. Instead, we have to make an inference. This inference involves sampling: instead of testing every member of the population, we test a small portion of the population—a sample— and infer from its properties to the properties of the whole. It's a simple inductive argument:

The sample has property X.

∴ The general population has property X.

The argument is inductive: the premise does not guarantee the truth of the conclusion; it merely makes it more probable. As was the case in hypothesis testing, we can be precise about the probabilities involved, and our probabilities come from the good-old bell curve.

Let's take a simple example.<sup>35</sup> Suppose we were trying to discover the percentage of men in the general population; we survey 100 people, and it turns out there are 55 men in our sample. So, the proportion of men in our sample is .55. We're trying to make an inference from this premise to a conclusion about the proportion of men in the general population. What's the probability that the proportion of men in the general population is .55? This isn't exactly the question we want to answer in these sorts of cases, though. Rather, we ask, what's the probability that the true proportion of men in the general population is in some range on either side of .55? We can give a precise answer to this question; the answer depends on the size of the range you're considering in a familiar way.

32. Except for those known to be at risk, who should start earlier.

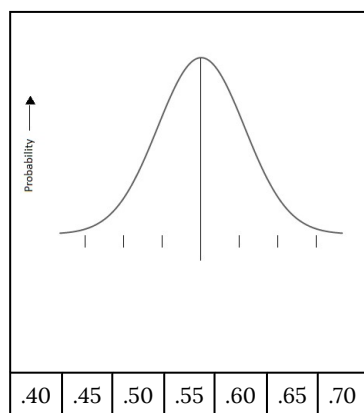
33. Especially perverse are the cases in which the radiation treatment itself causes cancer in a patient who didn't have to be treated to begin with.

34. PC Gøtzsche and KJ Jørgensen, 2013, Cochrane Database of Systematic Reviews (6), CD001877.pub5

35. I am indebted for this example in particular (and for much background on the presentation of statistical reasoning in general) to John Norton, 1998, *How Science Works*, New York: McGraw-Hill, pp. 12.14 – 12.15.

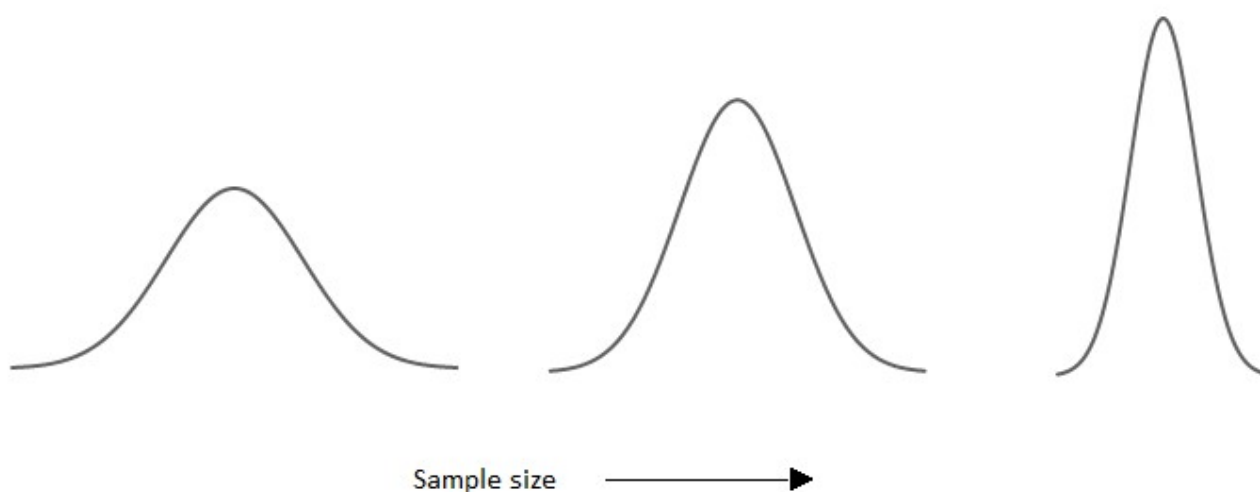
Given that our sample's proportion of men is .55, it is relatively more likely that the true proportion in the general population is close to that number, less likely that it's far away. For example, it's more likely, given the result of our survey, that in fact 50% of the population is men than it is that only 45% are men. And it's still less likely that only 40% are men. The same pattern holds in the opposite direction: it's more likely that the true percentage of men is 60% than 65%. Generally speaking, the further away from our survey results we go, the less probable it is that we have the true value for the general population. The drop off in probabilities described takes the form of a bell curve:

Proportion of Men in the Population



The standard deviation of .05 is a function of our sample size of 100.<sup>36</sup> We can use the usual confidence intervals—again, with 2 standard deviations, 95.4% being standard practice—to interpret the findings of our survey: we're pretty sure—to the tune of 95%—that the general population is between 45% and 65% male.

That's a pretty wide range. Our result is not that impressive (especially considering the fact that we know the actual number is very close to 50%). But that's the best we can do given the limitations of our survey. The main limitation, of course, was the size of our sample: 100 people just isn't very many. We could narrow the range within which we're 95% confident if we increased our sample size; doing so would likely (though not certainly) give us a proportion in our sample closer to the true value of (approximately) .5. The relationship between the sample size and the width of the confidence intervals is a purely mathematical one. As sample size goes up, standard deviation goes down—the curve narrows:



36. And the mean (our result of .55). The mathematical details of the calculation needn't detain us.

The pattern of reasoning on display in our toy example is the same as that used in sampling generally. Perhaps the most familiar instances of sampling in everyday life are public opinion surveys. Rather than trying to determine the proportion of people in the general population who are men (not a real mystery), opinion pollsters try to determine the proportion of a given population who, say, intend to vote for a certain candidate, or approve of the job the president is doing, or believe in Bigfoot. Pollsters survey a sample of people on the question at hand, and end up with a result: 29% of Americans believe in Bigfoot, for example.<sup>37</sup> But the headline number, as we have seen, doesn't tell the whole story. 29% of the sample (in this case, about 1,000 Americans) reported believing in Bigfoot; it doesn't follow with certainty that 29% of the general population (all Americans) have that belief. Rather, the pollsters have some degree of confidence (again, 95% is standard) that the actual percentage of Americans who believe in Bigfoot is in some range around 29%. You may have heard the "margin of error" mentioned in connection with such surveys. This phrase refers to the very range we're talking about. In the survey about Bigfoot, the margin of error is 3%. That's the distance from the mean (the 29% found in the sample) and the ends of the two standard-deviation confidence interval—the range in which we're 95% sure the true value lies. Again, this range is just a mathematical function of the sample size: if the sample size is around 100, the margin of error is about 10% (see the toy example above:  $2 \text{ SDs} = .10$ ); if the sample size is around 400, you get that down to 5%; at 600, you're down to 4%; at around 1,000, 3%; to get down to 2%, you need around 2,500 in the sample, and to get down to 1%, you need 10,000.<sup>38</sup> So the real upshot of the Bigfoot survey result is something like this: somewhere between 26% and 32% of Americans believe in Bigfoot, and we're 95% sure that's the correct range; or, to put it another way, we used a method for determining the true proportion of Americans who believe in Bigfoot that can be expected to determine a range in which the true value actually falls 95% of the time, and the range that resulted from our application of the method on this occasion was 26% – 32%.

That last sentence, we must admit, would make for a pretty lousy newspaper headline ("29% of Americans believe in Bigfoot!" is much sexier), but it's the most honest presentation of what the results of this kind of sampling exercise actually show. Sampling gives us a range, which will be wider or narrower depending on the size of the sample, and not even a guarantee that the actual value is within that range. That's the best we can do; these are inductive, not deductive, arguments.

Finally, on the topic of sampling, we should acknowledge that in actual practice, polling is hard. The mathematical relationships between sample size and margin of error/confidence that we've noted all hold in the abstract, but real-life polls can have errors that go beyond these theoretical limitations on their accuracy. As the 2016 U.S. presidential election—and the so-called "Brexit" vote in the United Kingdom that same year, and many, many other examples throughout the history of public opinion polling—showed us, polls can be systematically in error. The kinds of facts we've been stating—that with a sample size of 600, a poll has a margin of error of 4% at the 95% confidence level—hold only on the assumption that there's a systematic relationship between the sample and the general population it's meant to represent; namely, that the sample is *representative*. A representative sample mirrors the general population; in the case of people, this means that the sample and the general population have the same demographic make-up—same percentage of old people and young people, white people and people of color, rich people and poor people, etc., etc. Polls whose samples are not representative are likely to misrepresent the feature of the population they're trying to capture. Suppose I wanted to find out what percentage of the U.S. population thinks favorably of Donald Trump. If I asked 1,000 people in, say, rural Oklahoma, I'd get one result; if I asked 1,000 people in midtown Manhattan, I'd

37. Here's an actual survey with that result: [http://angusreidglobal.com/wp-content/uploads/2012/03/2012.03.04\\_Myths.pdf](http://angusreidglobal.com/wp-content/uploads/2012/03/2012.03.04_Myths.pdf)

38. Interesting mathematical fact: these relationships hold no matter how big the general population from which you're sampling (as long as it's above a certain threshold). It could be the size of the population of Wisconsin or the population of China: if your sample is 600 Wisconsinites, your margin of error is 4%; if it's 600 Chinese people, it's still 4%. This is counterintuitive, but true—at least, in the abstract. We're omitting the very serious difficulty that arises in actual polling (which we will discuss anon): finding the right 600 Wisconsinites or Chinese people to make your survey reliable; China will present more difficulty than Wisconsin.

get a much different result. Neither of those two samples is representative of the population of the United States as a whole. To get such a sample, I'd have to be much more careful about whom I surveyed. A famous example from the history of public polling illustrates the difficulties here rather starkly: in the 1936 U.S. presidential election, the contenders were Republican Alf Landon of Kansas, and the incumbent President Franklin D. Roosevelt. A (now-defunct) magazine, *Literary Digest* conducted a poll with 2.4 million (!) participants, and predicted that Landon would win in a landslide. Instead, he lost in a landslide; FDR won the second of his four presidential elections. What went wrong? With a sample size so large, the margin of error would be tiny. The problem was that their sample was not representative of the American population. They chose participants randomly from three sources: (a) their list of subscribers; (b) car registration forms; and (c) telephone listings. The problem with this selection procedure is that all three groups tended to be wealthier than average. This was 1936, during the depths of the Great Depression. Most people didn't have enough disposable income to subscribe to magazines, let alone have telephones or own cars. The survey therefore over-sampled Republican voters and got a skewed results. Even a large and seemingly random sample can lead one astray. This is what makes polling so difficult: finding representative samples is hard.<sup>39</sup>

Other practical difficulties with polling are worth noting. First, the way your polling question is worded can make a big difference in the results you get. As we discussed in Chapter 2, the framing of an issue—the words used to specify a particular policy or position—can have a dramatic effect on how a relatively uninformed person will feel about it. If you wanted to know the American public's opinion on whether or not it's a good idea to tax the transfer of wealth to the heirs of people whose holdings are more than \$5.5 million or so, you'd get one set of responses if you referred to the policy as an "estate tax", a different set of responses if you referred to it as an "inheritance tax", and a still different set if you called it the "death tax". A poll of Tennessee residents found that 85% opposed "Obamacare", while only 16% opposed "Insure Tennessee" (they're the same thing, of course).<sup>40</sup> Even slight changes in the wording of questions can alter the results of an opinion poll. This is why the polling firm Gallup hasn't changed the wording of its presidential-approval question since the 1930s. They always ask: "Do you approve or disapprove of the way [name of president] is handling his job as President?" A deviation from this standard wording can produce different results. The polling firm Ipsos found that its polls were more favorable than others' for the president. They traced the discrepancy to the different way they worded their question, giving an additional option: "Do you approve, disapprove, or have mixed feelings about the way Barack Obama is handling his job as president?"<sup>41</sup> A conjecture: Obama's approval rating would go down if pollsters included his middle name (Hussein) when asking the question. Small changes can make a big difference.

Another difficulty with polling is that some questions are harder to get reliable data about than others, simply because they involve topics about which people tend to be untruthful. Asking someone whether he approves of the job the president is doing is one thing; asking him whether or not he's ever cheated on his taxes, say, is quite another. He's probably not shy about sharing his opinion on the former question; he'll be much more reluctant to be truthful on the latter (assuming he's ever fudged things on his tax returns). There are lots of things it would be difficult to discover for this reason: how often people floss, how much they drink, whether or not they exercise, their sexual habits, and so on. Sometimes this reluctance to share the truth about oneself is quite consequential: some experts think that the

39. It's even harder than this paragraph makes it out to be. It's usually impossible for a sample—the people you've talked to on the phone about the president or whatever—to mirror the demographics of the population exactly. So pollsters have to weight the responses of certain members of their sample more than others to make up for these discrepancies. This is more art than science. Different pollsters, presented with the exact same data, will make different choices about how to weight things, and will end up reporting different results. See this fascinating piece for an example:

[http://www.nytimes.com/interactive/2016/09/20/upshot/the-errorthe pollingworldrarelytalksabout.html?\\_r=0](http://www.nytimes.com/interactive/2016/09/20/upshot/the-errorthe pollingworldrarelytalksabout.html?_r=0)

40. Source: <http://www.nbcnews.com/politics/elections/rebuke-tennesseegovernorkochgroupshowsitspowern301031-overnorkochgroupshowsitspowern301031>

41. <http://spotlight.ipsos-na.com/index.php/news/ispresidentobamaupordowntheeffectofquestionwordingonlevelsofpresidentialsupport/>

reason polls failed to predict the election of Donald Trump as president of the United States in 2016 was that some of his supporters were “shy”—unwilling to admit that they supported the controversial candidate.<sup>42</sup> They had no such qualms in the voting booth, however.

Finally, who’s asking the question—and the context in which it’s asked—can make a big difference. People may be more willing to answer questions in the relative anonymity of an online poll, slightly less willing in the somewhat more personal context of a telephone call, and still less forthcoming in a face-to-face interview. Pollsters use all of these methods to gather data, and the results vary accordingly. Of course, these factors become especially relevant when the question being polled is a sensitive one, or something about which people tend not to be honest or forthcoming. To take an example: the best way to discover how often people truly floss is probably with an anonymous online poll. People would probably be more likely to lie about that over the phone, and still more likely to do so in a face-to-face conversation. The absolute worst source of data on that question, perversely, would probably be from the people who most frequently ask it: dentists and dental hygienists. Every time you go in for a cleaning, they ask you how often you brush and floss; and if you’re like most people, you lie, exaggerating the assiduity with which you attend to your dental-health maintenance (“I brush after every meal and floss twice a day, honest.”).

As was the case with hypothesis testing, the logic of statistical sampling is relatively clear. Things get murky, again, when straightforward abstract methods confront the confounding factors involved in real-life application.

## EXERCISES

1. I and a bunch of my friends are getting ready to play a rousing game of “army men”. Together, we have 110 of the little plastic toy soldiers—enough for quite a battle. However, some of us have more soldiers than others. Will, Brian and I each have 25; Roger and Joe have 11 each; Dan has 4; John and Herb each have 3; Mike, Jamie, and Dennis have only 1 each.

(a) What is the mean number of army men held? What’s the median?

(b) Jamie, for example, is perhaps understandably disgruntled about the distribution; I, on the other hand, am satisfied with the arrangement. In defending our positions, each of us might refer to the “average person” and the number of army men he has. Which sense of ‘average’—mean or median—should Jamie use to gain a rhetorical advantage? Which should I use?

2. Consider cats and dogs—the domesticated kind, pets (tigers don’t count). Suppose I produced a histogram for a very large number of pet cats based on their weight, and did the same for pet dogs. Which distribution would have the larger standard deviation?

3. Men’s heights are normally distributed, with a mean of about 70 inches and a standard deviation of about 3 inches. 68.3% of men fall within what range of heights? Where do 95.4% of them fall? 99.7%? My father-in-law was 76 inches tall. What percentage of men were taller than he was?

4. Women, on average, have lower hematocrit levels than men. The mean for healthy women is 42%, with a standard deviation of 3%. Suppose we want to test the null hypothesis that Alice is healthy. What are the hematocrit readings above which and below which Alice’s test result would be considered significant at the .05 level?

5. Among healthy people, the mean (fasting) blood glucose level is 90 mg/dL, with a standard deviation of 9 mg/dL. What are the levels at the high and low end of the 95.4% confidence interval? Recently, I had my blood tested and got a result of 100 mg/dL. Is this result significant at the .05 level? My result was flagged as being potentially indicative of my being “pre-diabetic” (high blood glucose is a marker for diabetes). My doctor said this is a new standard, since

42. See here, for example: [https://www.washingtonpost.com/news/monkey-cage/wp/2016/12/13/whythepollsmisssedin2016wasitshytrumpsupportersafterall/?utm\\_term=.f20212063a9c](https://www.washingtonpost.com/news/monkey-cage/wp/2016/12/13/whythepollsmisssedin2016wasitshytrumpsupportersafterall/?utm_term=.f20212063a9c)

diabetes is on the rise lately, but I shouldn't worry because I wasn't overweight and was otherwise healthy. Compared to a testing regime that only flags patients outside the two standard-deviation confidence interval, does this new practice of flagging results at 100 mg/dL increase or decrease the sensitivity of the diabetes screening? Does it increase or decrease its specificity?

6. A stroke is when blood fails to reach a part of the brain because of an obstruction of a blood vessel. Often the obstruction is due to atherosclerosis—a hardening/narrowing of the arteries from plaque buildup. Strokes can be really bad, so it would be nice to predict them. Recent research has sought for a potentially predictive biomarker, and one study found that among stroke victims there was an unusually high level of an enzyme called myeloperoxidase: the mean was 583 pmol/L, with a standard deviation of 48 pmol/L.<sup>43</sup> Suppose we wanted to devise a screening test on the basis of this data. To guarantee that we caught 99.7% of potential stroke victims, what range of myeloperoxidase levels should get a “positive” test result? If the mean level of myeloperoxidase among healthy people is 425 pmol/L, with a standard deviation of 36 pmol/L, approximately what percentage of healthy people will get a positive result from our proposed screening test?

7. I survey a sample of 1,000 Americans (assume it's representative) and 43% of them report that they believe God created human beings in their present form less than 10,000 years ago.<sup>44</sup> At the 95% confidence level, what is the range within which the true percentage probably lies?

8. Volunteer members of Mothers Against Drunk Driving conducted a door-to-door survey in a college dormitory on a Saturday night, and discovered that students drink an average of two alcoholic beverages per week. What are some reasons to doubt the results of this survey?

## V. How to Lie with Statistics<sup>45</sup>

The basic grounding in fundamental statistical concepts and techniques provided in the last section gives us the ability to understand and analyze statistical arguments. Since real-life examples of such arguments are so often manipulative and misleading, our aim in this section is to build on the foundation of the last by examining some of the most common statistical fallacies—the bad arguments and deceptive techniques used to try to bamboozle us with numbers.

### *Impressive Numbers without Context*

I'm considering buying a new brand of shampoo. The one I'm looking at promises “85% more body”. That sounds great to me (I'm pretty bald; I can use all the extra body I can get). But before I make my purchase, maybe I should consider the fact that the shampoo bottle doesn't answer this simple follow-up question: 85% more body *than what*? The bottle does mention that the formulation inside is “new and improved”. So maybe it's 85% more body than the unimproved shampoo? Or possibly they mean that their shampoo gives hair 85% more body than their competitors'. Which competitor, though? The one that does the best at giving hair more body? The one that does the worst? The average of all the competing brands? Or maybe it's 85% more body than something else entirely. I once had a high school teacher who advised me to massage my scalp for 10 minutes every day to prevent baldness (I didn't take the suggestion; maybe I should have). Perhaps this shampoo produces 85% more body than daily 10-minute massages. Or

43. See this study: <https://www.ncbi.nlm.nih.gov/pubmed/21180247>

44. See this survey: <http://www.gallup.com/poll/27847/Majority-RepublicansDoubtTheoryEvolution.aspx>  
-oubtTheoryEvolution.aspx

45. The title of this section, a lot of the topics it discusses, and even some of the examples it uses, are taken from Huff 1954.

maybe it's 85% more body than never washing your hair at all. And just what is "body" anyway? How is it quantified and measured? Did they take high-precision calipers and systematically gauge the widths of hairs? Or is it more a function of coverage—hairs per square inch of scalp surface area?

The sad fact is, answers to these questions are not forthcoming. The claim that the shampoo will give my hair 85% more body sounds impressive, but without some additional information for me to contextualize that claim, I have no idea what it means. This is a classic rhetorical technique: throw out a large number to impress your audience, without providing the context necessary for them to evaluate whether or not your claim is actually all that impressive. Usually, on closer examination, it isn't. Advertisers and politicians use this technique all the time.

In the spring of 2009, the economy was in really bad shape (the fallout from the financial crisis that began in the fall of the year before was still being felt; stock market indices didn't hit their bottom until March 2009, and the unemployment rate was still on the rise). Barack Obama, the newly inaugurated president at the time, wanted to send the message to the American people that he got it: households were cutting back on their spending because of the recession, and so the government would do the same thing.<sup>46</sup> After his first meeting with his cabinet (the Secretaries of Defense, State, Energy, etc.), he held a press conference in which he announced that he had ordered each of them to cut \$100 million from their agencies' budgets. He had a great line to go with the announcement: "\$100 million there, \$100 million here—pretty soon, even here in Washington, it adds up to real money." Funny. And impressive-sounding. \$100 million is a hell of a lot of money! At least, it's a hell of a lot of money to me. I've got—give me a second while I check—\$64 in my wallet right now. I wish I had \$100 million. But of course my personal finances are the wrong context in which to evaluate the president's announcement. He's talking about cutting from the federal budget; that's the context. How big is that? In 2009, it was a little more the \$3 trillion. There are fifteen departments that the members of the cabinet oversee. The cut Obama ordered amounted to \$1.5 billion, then. That's .05% of the federal budget. That number's not sounding as impressive now that we put it in the proper context.

2009 provides another example of this technique. Opponents of the Affordable Care Act ("Obamacare") complained about the length of the bill: they repeated over and over that it was 1,000 pages long. That complaint dovetailed nicely with their characterization of the law as a boondoggle and a government takeover of the healthcare system. 1,000 pages sure sounds like a lot of pages. This book comes in under 250 pages; imagine if it were 1,000! That would be up there with notoriously long books like *War and Peace*, *Les Miserable*, and *Infinite Jest*. It's long for a book, but is it a lot of pages for a piece of federal legislation? Well, it's big, but certainly not unprecedented. That year's stimulus bill was about the same length. President Bush's 2007 budget bill was just shy of 1,500 pages.<sup>47</sup> His No Child Left Behind bill clocks in at just shy of 700. The fact is, major pieces of legislation have a lot of pages. The Affordable Care Act was not especially unusual.

## *Misunderstanding Error*

As we discussed, built in to the logic of sampling is a margin of error. It is true of measurement generally that random error is unavoidable: whether you're measuring length, weight, velocity, or whatever, there are inherent limits to the precision and accuracy with which our instruments can measure things. Measurement errors are built in to the logic of scientific practice generally; they must be accounted for. Failure to do so—or intentionally ignoring error—can produce misleading reports of findings.

46. This sounds good, but it's bad macroeconomics. Most economists agree that during a downturn like that one, the government should borrow and spend more, not less, in order to stimulate the economy. The president knew this; he ushered a huge government spending bill through Congress (The American Reinvestment and Recovery Act) later that year.

47. This is a useful resource: [http://www.slate.com/articles/news\\_and\\_politics/explainer/2009/08/paper\\_weight.html](http://www.slate.com/articles/news_and_politics/explainer/2009/08/paper_weight.html)



This is particularly clear in the case of public opinion surveys. As we saw, the results of such polls are not the precise percentages that are often reported, but rather ranges of possible percentages (with those ranges only being reliable at the 95% confidence level, typically). And so to report the results of a survey, for example, as “29% of Americans believe in Bigfoot”, is a bit misleading since it leaves out the margin of error and the confidence level. A worse sin is committed (quite commonly) when comparisons between percentages are made and the margin of error is omitted. This is typical in politics, when the levels of support for two contenders for an office are being measured. A typical newspaper headline might report something like this: “Candidate Sue Surges into the Lead over Incumbent Bob in Latest Poll, 44% to 43%”. This is a sexy headline: it’s likely to sell papers (or, nowadays, generate clicks), both to (happy) Sue supporters and (alarmed) Bob supporters. But it’s misleading: it suggests a level of precision, a definitive result, that the data simply do not support. Let’s suppose that the margin of error for this hypothetical poll was 3%. What the survey results actually tell us, then, is that (at the 95% confidence level) the true level of support for Sue in the general population is somewhere between 41% and 47%, while the true level of support for Bob is somewhere between 40% and 46%. Those data are consistent with a Sue lead, to be sure; but they also allow for a commanding 46% to 41% lead for Bob. The best we can say is that it’s slightly more likely that Sue’s true level of support is higher than Bob’s (at least, we’re pretty sure; 95% confidence interval and all). When differences are smaller than the margin of error (really, twice the margin of error when comparing two numbers), they just don’t mean very much. That’s a fact that headline-writers typically ignore. This gives readers a misleading impression about the certainty with which the state of the race can be known.

Early in their training, scientists learn that they cannot report values that are smaller than the error attached to their measurements. If you weigh some substance, say, and then run an experiment in which it’s converted into a gas, you can plug your numbers into the ideal gas law and punch them into your calculator, but you’re not allowed to report all the numbers that show up after the decimal place. The number of so-called “significant digits” (or sometimes “figures”) you can use is constrained by the size of the error in your measurements. If you can only know the original weight to within .001 grams, for example, then even though the calculator spits out .4237645, you can only report a result using three significant digits—.424 after rounding.

The more significant digits you report, the more precise you imply your measurement is. This can have the rhetorical effect of making your audience easier to persuade. Precise numbers are impressive; they give people the impression that you really know what you’re talking about, that you’ve done some serious quantitative analytical work. Suppose I ask 1,000 college students how much sleep they got last night.<sup>48</sup> I add up all the numbers and divide by 1,000, and my calculator gives me 7.037 hours. If I went around telling people that I’d done a study that showed that the average college student gets 7.037 hours of sleep per night, they’d be pretty impressed: my research methods were so thorough that I can report sleep times down to the thousandths of an hour. They’ve probably got a mental picture of my laboratory, with elaborate equipment hooked up to college students in beds, measuring things like rapid eye movement and breathing patterns to determine the precise instants at which sleep begins and ends. But I have no such laboratory. I just asked a bunch of people. Ask yourself: how much sleep did you get last night? I got about 9 hours (it’s the weekend). The key word in that sentence is ‘about’. Could it have been a little bit more or less than 9 hours? Could it have been 9 hours and 15 minutes? 8 hours and 45 minutes? Sure. The error on any person’s report of how much they slept last night is bound to be something like a quarter of an hour. That means that I’m not entitled to those 37 thousandths of an hour that I reported from my little survey. The best I can do is say that the average college student gets about 7 hours of sleep per night, plus or minus 15 minutes or so. 7.037 is precise, but the precision of that figure is spurious (not genuine, false).

Ignoring the error attached to measurements can have profound real-life effects. Consider the 2000 U.S. presidential election. George W. Bush defeated Al Gore that year, and it all came down to the state of Florida, where the final margin of victory (after recounts were started, then stopped, then started again, then finally stopped by order of the Supreme Court of the United States) was 327 votes. There were about 6 million votes cast in Florida that year. The margin of 327 is about .005% of the total. Here’s the thing: counting votes is a measurement like any other; there is an error attached

48. This example inspired by Huff 1954, pp. 106 – 107.

to it. You may remember that in many Florida counties, they were using punch-card ballots, where voters indicate their preference by punching a hole through a perforated circle in the paper next to their candidate's name. Sometimes, the circular piece of paper—a so-called “chad”—doesn't get completely detached from the ballot, and when that ballot gets run through the vote-counting machine, the chad ends up covering the hole and a non-vote is mistakenly registered. Other types of vote-counting methods—even hand-counting<sup>49</sup>—have their own error. And whatever method is used, the error is going to be greater than the .005% margin that decided the election. As one prominent mathematician put it, “We're measuring bacteria with a yardstick.”<sup>50</sup> That is, the instrument we're using (counting, by machine or by hand) is too crude to measure the size of the thing we're interested in (the difference between Bush and Gore). He suggested they flip a coin to decide Florida. It's simply impossible to know who won that election.

In 2011, newly elected Wisconsin Governor Scott Walker, along with his allies in the state legislature, passed a budget bill that had the effect, among other things, of cutting the pay of public sector employees by a pretty significant amount. There was a lot of uproar; you may have seen the protests on the news. People who were against the bill made their case in various ways. One of the lines of attack was economic: depriving so many Wisconsin residents of so much money would damage the state's economy and cause job losses (state workers would spend less, which would hurt local businesses' bottom lines, which would cause them to lay off their employees). One newspaper story at the time quoted a professor of economics who claimed that the Governor's bill would cost the state 21,843 jobs.<sup>51</sup> Not 21, 844 jobs; it's not that bad. Only 21,843. This number sounds impressive; it's very precise. But of course that precision is spurious. Estimating the economic effects of public policy is an extremely uncertain business. I don't know what kind of model this economist was using to make his estimate, but whatever it was, it's impossible for its results to be reliable enough to report that many significant digits. My guess is that at best the 2 in 21,843 has any meaning at all.

### *Tricky Percentages*

Statistical arguments are full of percentages, and there are lots of ways you can fool people with them. The key to not being fooled by such figures, usually, is to keep in mind *what it's a percentage of*. Inappropriate, shifting, or strategically chosen numbers can give you misleading percentages.

When the numbers are very small, using percentages instead of fractions is misleading. Johns Hopkins Medical School, when it opened in 1893, was one of the few medical schools that allowed women to matriculate.<sup>52</sup> In those benighted times, people worried about women enrolling in schools with men for a variety of silly reasons. One of them was the fear that the impressionable young ladies would fall in love with their professors and marry them. Absurd, right? Well, maybe not: in the first class to enroll at the school, 33% of the women did indeed marry their professors! The sexists were apparently right. That figure sounds impressive, until you learn that the denominator is 3. Three women enrolled at Johns Hopkins that first year, and one of them married her anatomy professor. Using the percentage rather than the fraction exaggerates in a misleading way. Another made up example: I live in a relatively safe little town. If I saw a headline in my local newspaper that said “Armed Robberies are Up 100% over Last Year” I would be quite alarmed. That is, until I realized that last year there was one armed robbery in town, and this year there were two. That is a 100% increase, but using the percentage of such a small number is misleading.

You can fool people by changing the number you're taking a percentage of mid-stream. Suppose you're an employee at

49. It may be as high as 2% for hand-counting! See here: <https://www.sciencedaily.com/releases/2012/02/120202151713.htm>

50. John Paulos, “We're Measuring Bacteria with a Yardstick,” November 22, 2000, The New York Times.

51. Steven Verburg, “Study: Budget Could Hurt State's Economy,” March 20, 2011, Wisconsin State Journal.

52. Not because the school's administration was particularly enlightened. They could only open with the financial support of four wealthy women who made this a condition for their donations.

my aforementioned LogiCorp. You evaluate arguments for \$10.00 per hour. One day, I call all my employees together for a meeting. The economy has taken a turn for the worse, I announce, and we've got fewer arguments coming in for evaluation; business is slowing. I don't want to lay anybody off, though, so I suggest that we all share the pain: I'll cut everybody's pay by 20%; but when the economy picks back up, I'll make it up to you. So you agree to go along with this plan, and you suffer through a year of making a mere \$8.00 per hour evaluating arguments. But when the year is up, I call everybody together and announce that things have been improving and I'm ready to set things right: starting today, everybody gets a 20% raise. First a 20% cut, now a 20% raise; we're back to where we were, right? Wrong. I changed numbers midstream. When I cut your pay initially, I took twenty percent of \$10.00, which is a reduction of \$2.00. When I gave you a raise, I gave you twenty percent of your reduced pay rate of \$8.00 per hour. That's only \$1.60. Your final pay rate is a mere \$9.60 per hour.<sup>53</sup>

Often, people make a strategic decision about what number to take a percentage of, choosing the one that gives them a more impressive-sounding, rhetorically effective figure. Suppose I, as the CEO of LogiCorp, set an ambitious goal for the company over the next year: I propose that we increase our productivity from 800 arguments evaluated per day to 1,000 arguments per day. At the end of the year, we're evaluating 900 arguments per day. We didn't reach our goal, but we did make an improvement. In my annual report to investors, I proclaim that we were 90% successful. That sounds good; 90% is really close to 100%. But it's misleading. I chose to take a percentage of 1,000: 900 divided by 1,000 give us 90%. But is that the appropriate way to measure the degree to which we met the goal? I wanted to increase our production from 800 to 1,000; that is, I wanted a total increase of 200 arguments per day. How much of an increase did we actually get? We went from 800 up to 900; that's an increase of 100. Our goal was 200, but we only got up to 100. In other words, we only got to 50% of our goal. That doesn't sound as good.

Another case of strategic choices. Opponents of abortion rights might point out that 97% of gynecologists in the United States have had patients seek abortions. This creates the impression that there's an epidemic of abortion-seeking, that it happens regularly. Someone on the other side of the debate might point out that only 1.25% of women of childbearing age get an abortion each year. That's hardly an epidemic. Each of the participants in this debate has chosen a convenient number to take a percentage of. For the anti-abortion activist, that is the number of gynecologists. It's true that 97% have patients who seek abortions; only 14% of them actually perform the procedure, though. The 97% exaggerates the prevalence of abortion (to achieve a rhetorical effect). For the pro-choice activist, it is convenient to take a percentage of the total number of women of childbearing age. It's true that a tiny fraction of them get abortions in a given year; but we have to keep in mind that only a small percentage of those women are pregnant in a given year. As a matter of fact, among those that actually get pregnant, something like 17% have an abortion. The 1.25% minimizes the prevalence of abortion (again, to achieve a rhetorical effect).

### *The Base-Rate Fallacy*

The base rate is the frequency with which some kind of event occurs, or some kind of phenomenon is observed. When we ignore this information, or forget about it, we commit a fallacy and make mistakes in reasoning.

Most car accidents occur in broad daylight, at low speeds, and close to home. So does that mean I'm safer if I drive really fast, at night, in the rain, far away from my house? Of course not. Then why are there more accidents in the former conditions? The base rates: much more of our driving time is spent at low speeds, during the day, and close to home; relatively little of it is spent driving fast at night, in the rain and far from home.<sup>54</sup>

Consider a woman formerly known as Mary (she changed her name to Moon Flower). She's a committed pacifist, vegan, and environmentalist; she volunteers with Green Peace; her favorite exercise is yoga. Which is more probable: that she's

53. This example inspired by Huff 1954, pp. 110 - 111.

54. This example inspired by Huff 1954, pp. 77 - 79.

a best-selling author of new-age, alternative-medicine, self-help books—or that she's a waitress? If you answered that she's more likely to be a best-selling author of self-help books, you fell victim to the base-rate fallacy. Granted, Moon Flower fits the stereotype of the kind of person who would be the author of such books perfectly. Nevertheless, it's far more probable that a person with those characteristics would be a waitress than a best-selling author. Why? Base rates. There are *far, far (far!)* more waitresses in the world than best-selling authors (of new-age, alternative-medicine, self-help books). The base rate of waitressing is higher than that of best-selling authorship by many orders of magnitude.

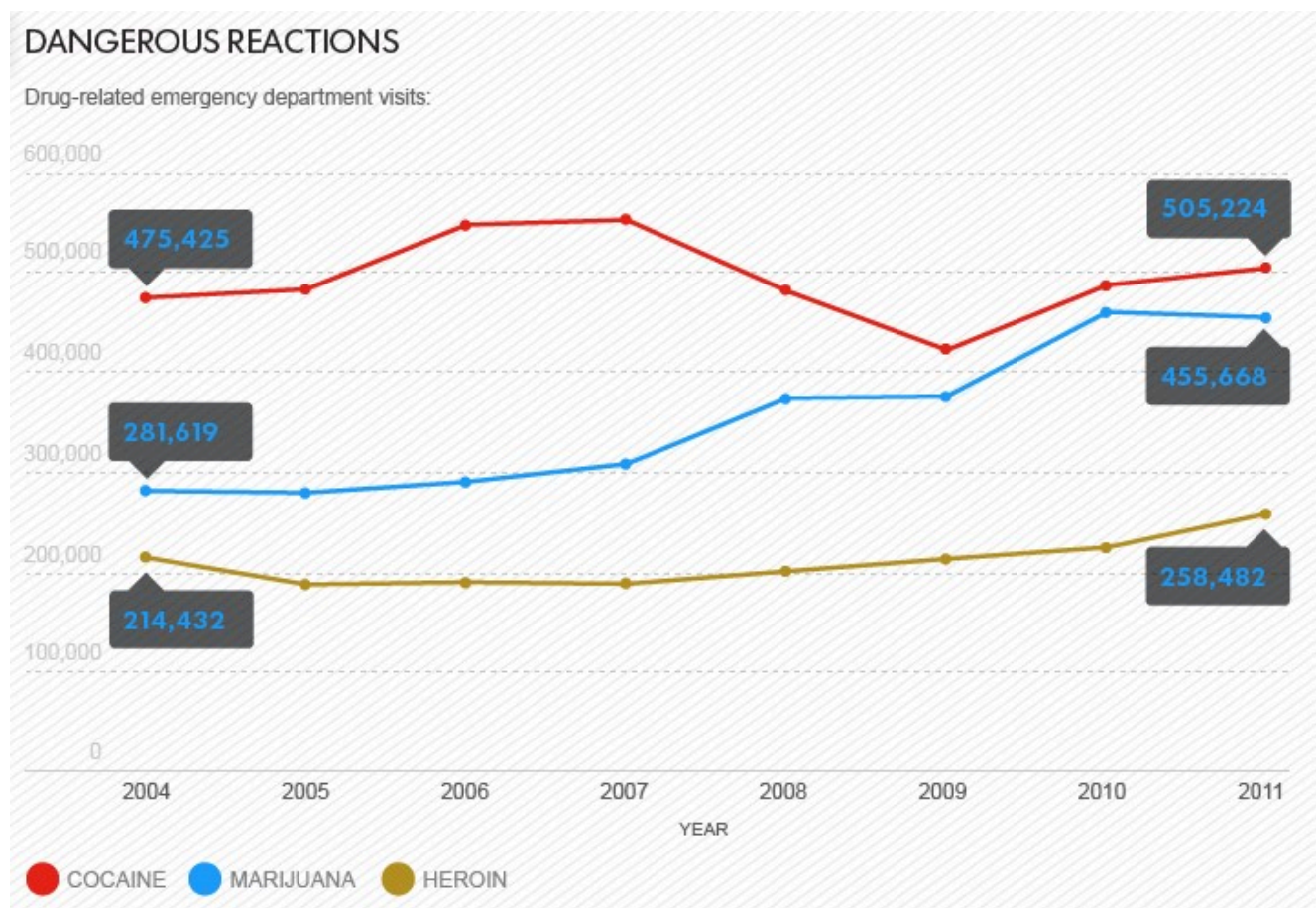
Suppose there's a medical screening test for a serious disease that is very accurate: it only produces false positives 1% of the time, and it only produces false negatives 1% of the time (it's highly sensitive and highly specific). The disease is serious, but rare: it only occurs in 1 out of every 100,000 people. Suppose you get screened for this disease and your result is positive; that is, you're flagged as possibly having the disease. Given what we know, what's the probability that you're actually sick? It's not 99%, despite the accuracy of the test. It's much lower. And I can prove it, using our old friend Bayes' Law. The key to seeing why the probability is much lower than 99%, as we shall see, is taking the base rate of the disease into account.

There are two hypotheses to consider: that you're sick (call it 'S') and that you're not sick (~ S). The evidence we have is a positive test result (P). We want to know the probability that you're sick, given this evidence:  $P(S | P)$ . Bayes' Law tells us how to calculate this:

$$P(S | P) = \frac{P(S) \times P(P | S)}{P(S) \times P(P | S) + P(\sim S) \times P(P | \sim S)}$$

The base rate of the sickness is the rate at which it occurs in the general population. It's rare: it only occurs in 1 out of 100,000 people. This number corresponds to the prior probability for the sickness in our formula— $P(S)$ . We have to multiply in the numerator by  $1/100,000$ ; this will have the effect of keeping down the probability of sickness, even given the positive test result. What about the other terms in our equation? ' $P(\sim S)$ ' just picks out the prior probability of not being sick; if  $P(S) = 1/100,000$ , then  $P(\sim S) = 99,999/100,000$ . ' $P(P | S)$ ' is the probability that you would get a positive test result, assuming you were in fact sick. We're told that the test is very accurate: it only tells sick people that they're healthy 1% of the time (1% rate of false negatives); so the probability that a sick person would get a positive test result is 99%— $P(P | S) = .99$ . ' $P(P | \sim S)$ ' is the probability that you'd get a positive result if you weren't sick. That's the rate of false positives, which is 1%— $P(P | \sim S) = .01$ . Plugging these numbers into the formula, we get the result that  $P(S | P) = .000999$ . That's right, given a positive result from this very-accurate screening test, you're probability of being sick is just under  $1/10,000$ . The test is accurate, but the disease is so rare (its base rate is so low) that your chances of being sick are still very low even after a positive result.

Sometimes people will ignore base rates on purpose to try to fool you. Did you know that marijuana is more dangerous than heroin? Neither did I. But look at this chart:



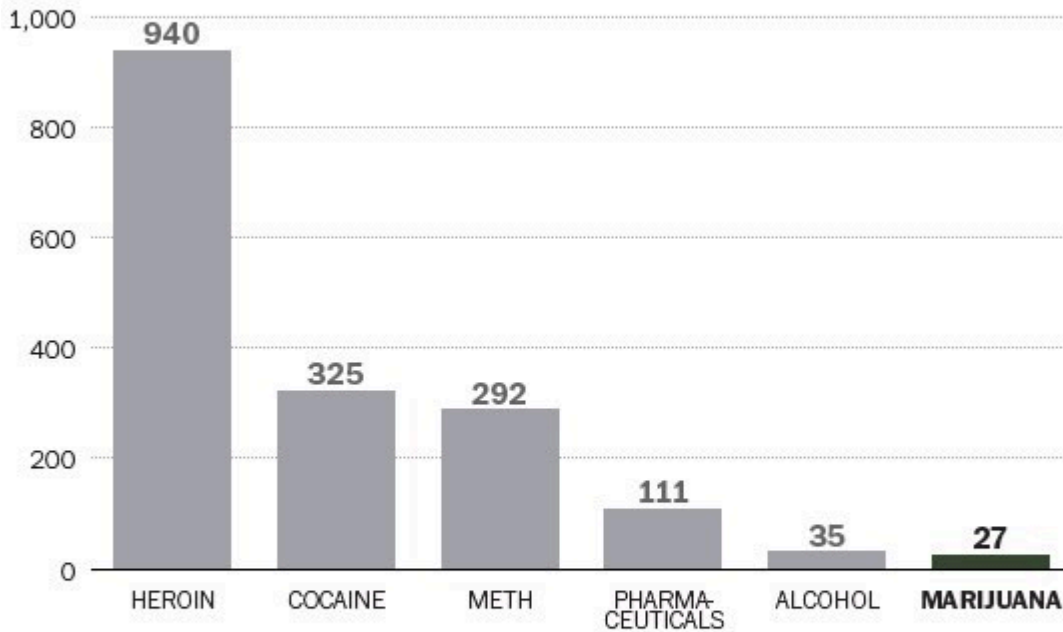
That graphic published in a story in USA Today under the headline “Marijuana poses more risks than many realize.”<sup>55</sup> The chart/headline combo create an alarming impression: if so many more people are going to the emergency room because of marijuana, it must be more dangerous than I realized. Look at that: more than twice as many emergency room visits for pot than heroin; it’s almost as bad as cocaine! Or maybe not. What this chart ignores is the base rates of marijuana-, cocaine-, and heroin-use in the population. Far (far!) more people use marijuana than use heroin or cocaine. A truer measure of the relative dangers of the various drugs would be the number of emergency room visits *per user*. That gives you a far different chart:<sup>56</sup>

55. Liz Szabo, “Marijuana poses more risks than many realize,” July 27, 2014, USA Today. <http://www.usatoday.com/story/news/nation/2014/07/27/risks-ofmarijuana/10386699/?sf29269095=1>

56. From German Lopez, “Marijuana sends more people to the ER than heroin. But that’s not the whole story.” August 2, 2014, Vox.com. <http://www.vox.com/2014/8/2/5960307/marijuana-legalizationheroinUSAToday>

# On a per-user basis, marijuana causes fewer ER trips than alcohol and other drugs

Emergency room visits per 1,000 users, 2010



Notes: "Users" are those who report using the substance at any time in the past month. Emergency room visits are counted if the patient was treated for a condition induced by or related to the given substance. Multiple substances can be reported in one ER visit.

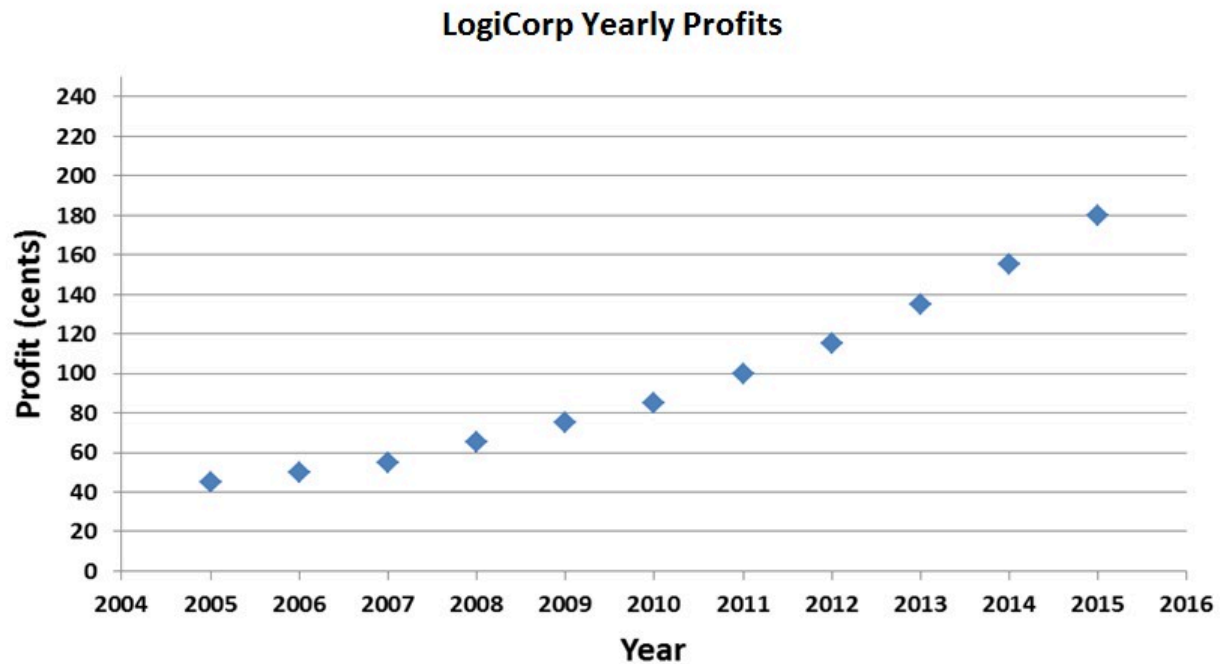
Sources: 2010 National Survey on Drug Use and Health; Drug Abuse Warning Network 2010; "Alcohol-related emergency department visits and hospitalizations" (National Institutes of Health)

WASHINGTONPOST.COM/WONKBLOG

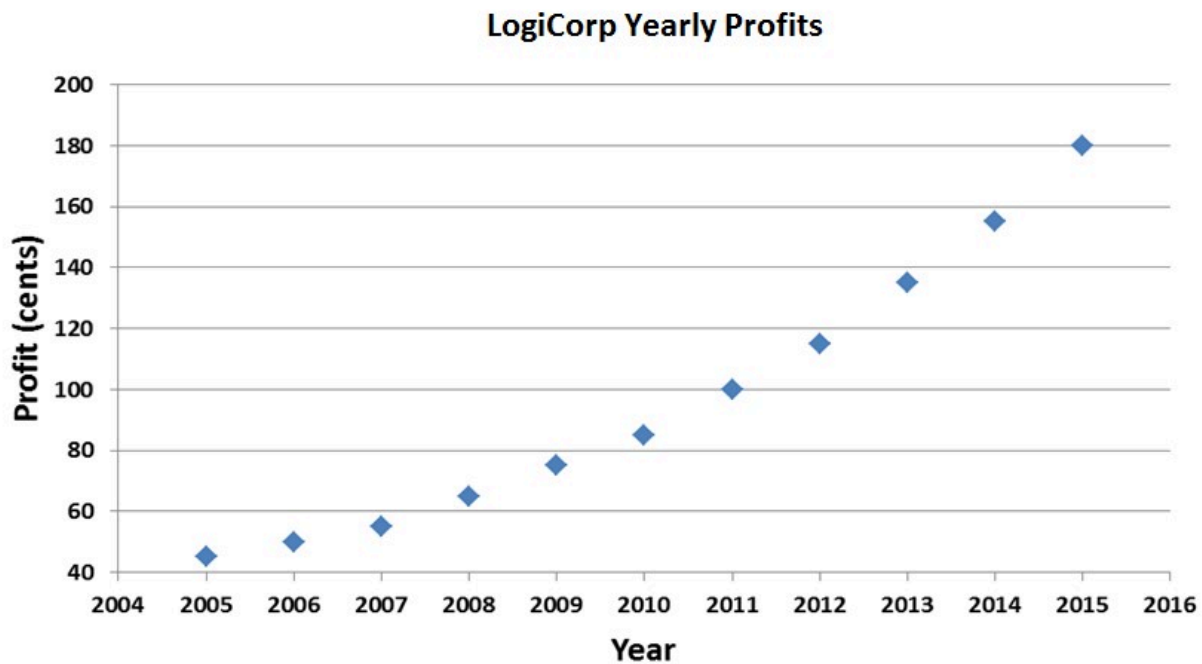
## *Lying with Pictures*

Speaking of charts, they are another tool that can be used (abused) to make dubious statistical arguments. We often use charts and other pictures to graphically convey quantitative information. But we must take special care that our pictures accurately depict that information. There are all sorts of ways in which graphical presentations of data can distort the actual state of affairs and mislead our audience.

Consider, once again, my fictional company, LogiCorp. Business has been improving lately, and I'm looking to get some outside investors so I can grow even more quickly. So I decide to go on that TV show *Shark Tank*. You know, the one with Mark Cuban and panel of other rich people, where you make a presentation to them and they decide whether or not your idea is worth investing in. Anyway, I need to plan a persuasive presentation to convince one of the sharks to give me a whole bunch of money for LogiCorp. I'm going to use a graph to impress them with company's potential for future growth. Here's a graph of my profits over the last decade:

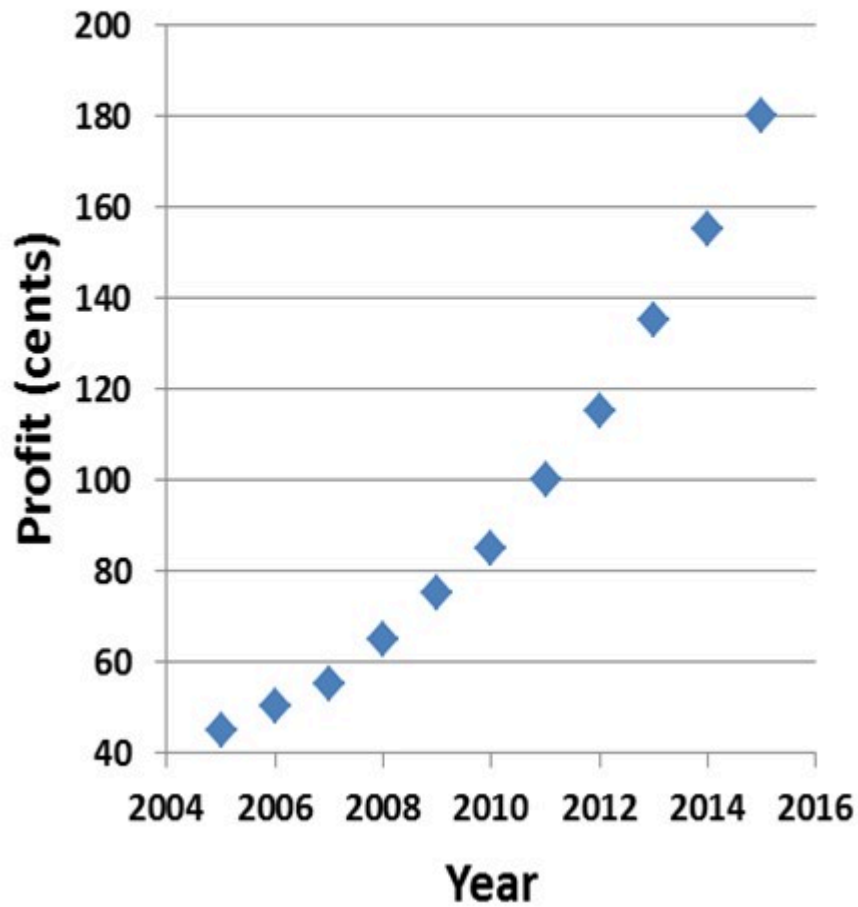


Not bad. But not great, either. The positive trend in profits is clearly visible, but it would be nice if I could make it look a little more dramatic. I'll just tweak things a bit:



Better. All I did was adjust the y-axis. No reason it has to go all the way down to zero and up to 240. Now the upward slope is accentuated; it looks like LogiCorp is growing more quickly. But I think I can do even better. Why does the x-axis have to be so long? If I compressed the graph horizontally, my curve would slope up even more dramatically:

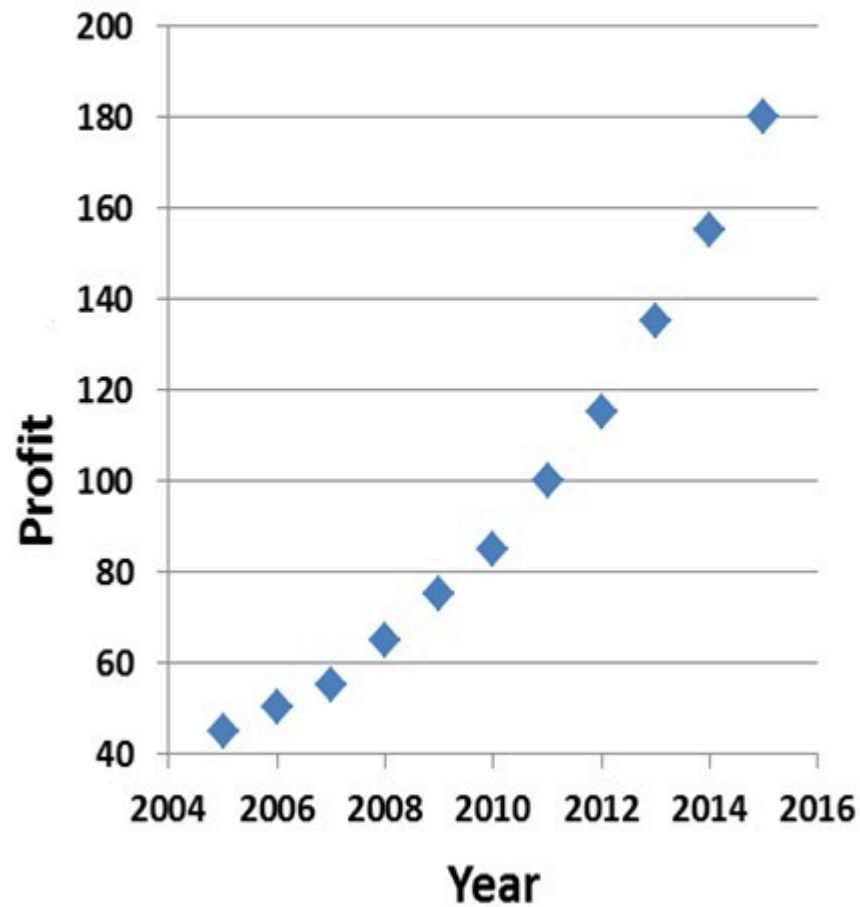
## LogiCorp Yearly Profits



Now that's explosive growth! The sharks are gonna love this. Well, that is, as long as they don't look too closely at the chart. Profits on the order of \$1.80 per year aren't going to impress a billionaire like Mark Cuban. But I can fix that:

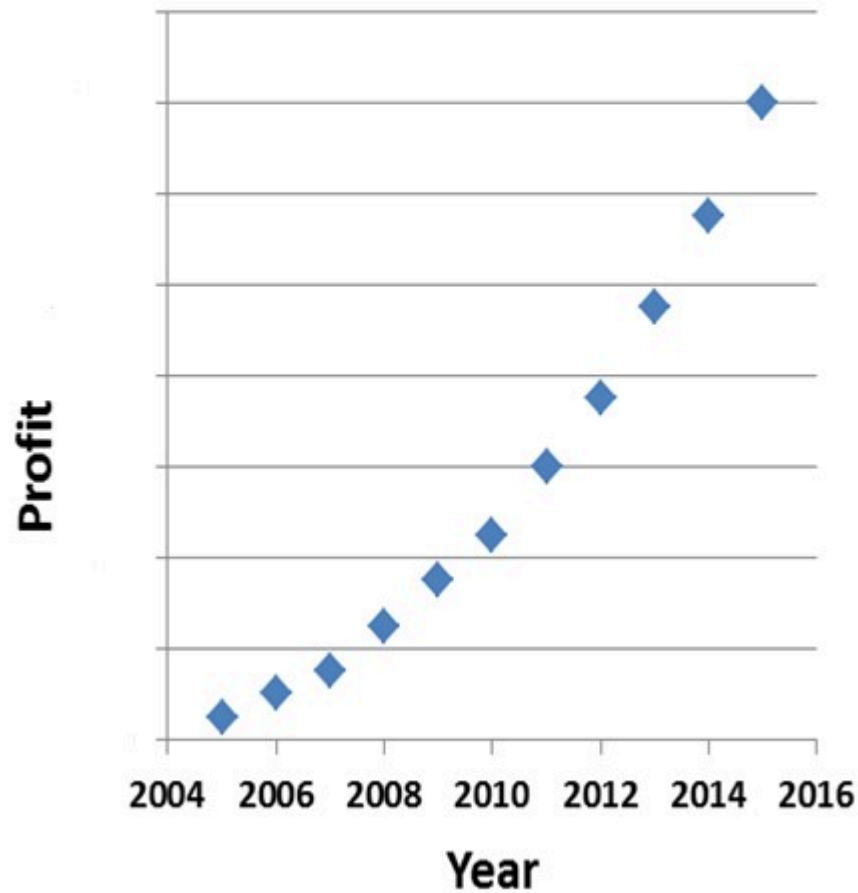


## LogiCorp Yearly Profits



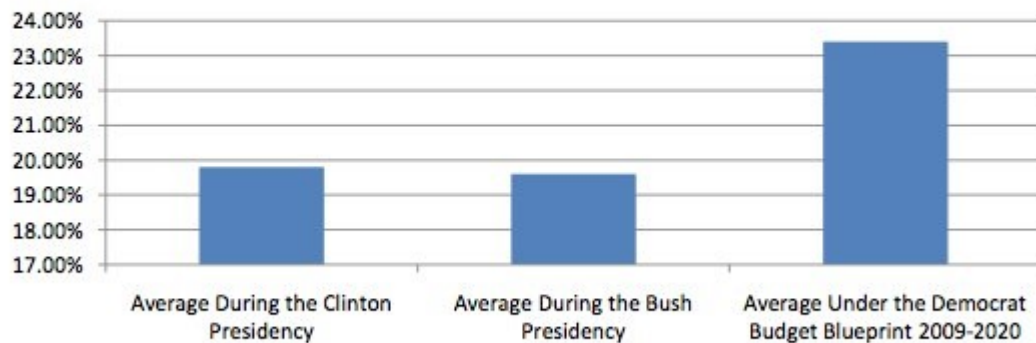
There. For all those sharks know, profits are measure in the millions of dollars. Of course, for all my manipulations, they can still see that profits have increased 400% over the decade. That's pretty good, of course, but maybe I can leave a little room for them to mentally fill in more impressive numbers:

**LogiCorp Yearly Profits**



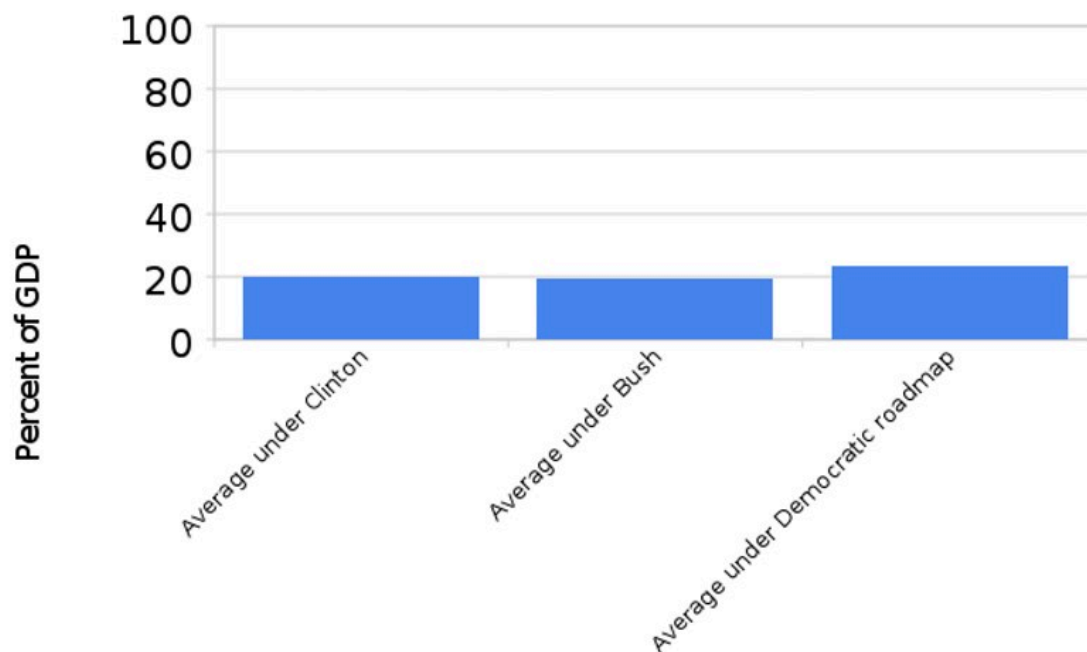
That's the one. Soaring profits, and it looks like they started close to zero and went up to—well, we can't really tell. Maybe those horizontal lines go up in increments of 100, or 1,000. LogiCorp's profits could be unimaginably high. People manipulate the y-axis of charts for rhetorical effect all the time. In their "Pledge to America" document of 2010, the Republican Party promised to pursue various policy priorities if they were able to achieve a majority in the House of Representatives (which they did). They included the following chart in that diagram to illustrate that government spending was out of control:

## Federal Spending as a Share of the Economy



Writing for *New Republic*, Alexander Hart pointed out that the Republicans' graph, by starting the y-axis at 17% and only going up to 24%, exaggerates the magnitude of the increase. That bar on the right is more than twice as big as the other two, but federal spending hadn't doubled. He produced the following alternative presentation of the data<sup>57</sup>:

### Honest graph

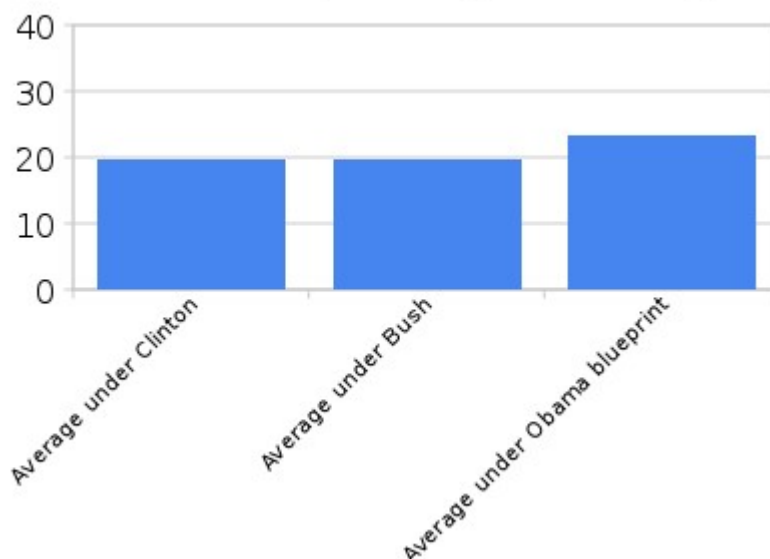


Writing for *The Washington Post*, liberal blogger Ezra Klein passed along the original graph and the more "honest" one. Many of his commenters (including your humble author) pointed out that the new graph was an over-correction of the

57. Alexander Hart, "Lying With Graphs, Republican Style (Now Featuring 50% More Graphs)," December 22, 2010, *New Republic*. <https://newrepublic.com/article/77893/lying-graphsrepublicanstyle>

first: it minimizes the change in spending by taking the y-axis all the way up to 100. He produced a final graph that's probably the best way to present the spending data<sup>58</sup>:

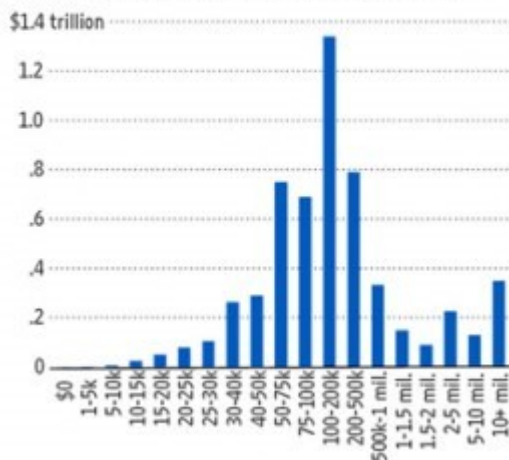
**Government size as percentage of economy**



One can make mischief on the x-axis, too. In an April 2011 editorial entitled “Where the Tax Money Is”, *The Wall Street Journal* made the case that President Obama’s proposal to raise taxes on the rich was a bad idea.<sup>59</sup> If he was really serious about raising revenue, he would have to raise taxes on the middle class, since that’s where most of the money is. To back up that claim, they produced this graph:

**The Middle Class Tax Target**

The amount of total taxable income (left scale) for all filers by adjusted gross income level for 2008



Source: IRS

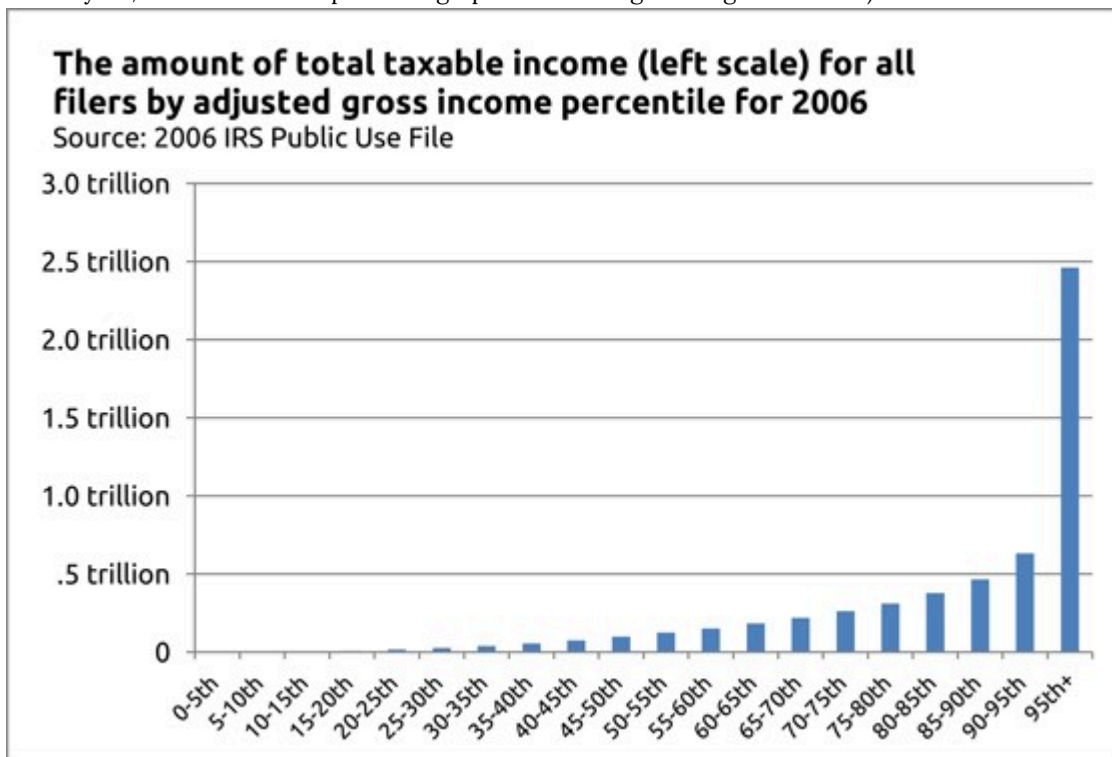
This one is subtle. What they present has the appearance of a histogram, but it breaks one of the rules for such charts: each of the bars has to represent the same portion of the population. That’s not even close to the case here. To get their tall bars in the middle of the income distribution, the *Journal’s* editorial board groups together incomes between \$50

58. Ezra Klein, “Lies, damn lies, and the ‘Y’ axis,” September 23, 2010, The Washington Post.

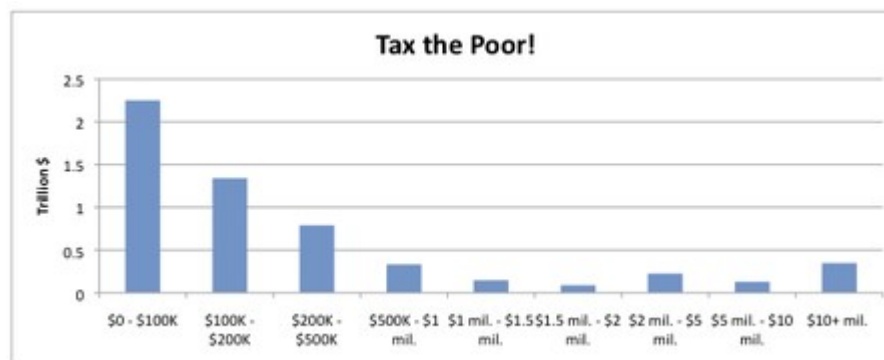
[http://voices.washingtonpost.com/ezra-klein/2010/09/lies\\_damn\\_lies\\_and\\_the\\_y\\_axis.html](http://voices.washingtonpost.com/ezra-klein/2010/09/lies_damn_lies_and_the_y_axis.html)

59. See here: <http://www.wsj.com/articles/SB10001424052748704621304576267113524583554>

and \$75 thousand, \$75 and \$100 thousand, then \$100 and \$200 thousand, and so on. There are far (far!) more people (or probably households; that's how these data are usually reported) in those income ranges than there are in, say, the range between \$20 and \$25 thousand, or \$5 to \$10 million—and yet those ranges get their own bars, too. That's just not how histograms work. Each bar in an income distribution chart would have to contain the same number of people (or households). When you produce such a histogram, you see what the distribution really looks like (these data are from a different tax year, but the basic shape of the graph didn't change during the interim):



Using *The Wall Street Journal's* method of generating histograms—where each bar can represent any number of different households—you can “prove” anything you like. It's not the rich or even the middle class we should go after if we really want to raise revenue; it's the poor. That's where the money is:



There are other ways besides charts and graphs to visually present quantitative information: pictograms. There's a sophisticated and rule-based method for representing statistical information using such pictures. It was pioneered in the 1920s by the Austrian philosopher Otto Neurath, and was originally called the Vienna Method of Pictorial Statistics (*Wiener Methode der Bildstatistik*); eventually it came to be known as Isotype (International System of TYpographic Picture Education)<sup>60</sup>. The principles of Neurath's system were such as to prevent the misrepresentation of data with

60. See here: [https://en.wikipedia.org/wiki/Isotype\\_\(picture\\_language\)](https://en.wikipedia.org/wiki/Isotype_(picture_language))

pictograms. Perhaps the most important rule is that greater quantities are to be represented not by larger pictures, but by greater numbers of same-sized pictures. So, for instance, if I wanted to represent the fact that domestic oil production in the United States has doubled over the past several years, I could use the following depiction<sup>61</sup>:



*THEN*

*NOW*

It would be misleading to flout Neurath's principles and instead represent the increase with a larger barrel:



*THEN*

*NOW*

All I did was double the size of the image. But I doubled it in both dimensions: it's both twice as wide and twice as tall. Moreover, since oil barrels are three dimensional objects, I've also depicted a barrel on the right that's twice as deep. The important thing about oil barrels is how much oil they can hold—their volume. By doubling the barrel in all three dimensions, I've depicted a barrel on the right that can hold 8 times as much oil as the one on the left. What I'm showing isn't a doubling of oil production; it's an eight-fold increase.

61. I've been using this example in class for years, and something tells me I got it from somebody else's book, but I've looked through all the books on my shelves and can't find it. So maybe I made it up myself. But if I didn't, this footnote acknowledges whoever did. (If you're that person, let me know!)

Alas, people break Neurath's rules all the time, and end up (intentionally or not) exaggerating the phenomena they're trying to depict. Matthew Yglesias, writing in *Architecture* magazine, made the point that the housing "bubble" that reached full inflation in 2006 (when lots of homes were built) was not all that unusual. If you look at recent history, you see similar cycles of boom and bust, with periods of lots of building followed by periods of relatively little. The magazine produced a graphic to present the data on home construction, and Yglesias made a point to post it on his blog at Slate.com because he thought it was illustrative.<sup>62</sup> Here's the graphic:



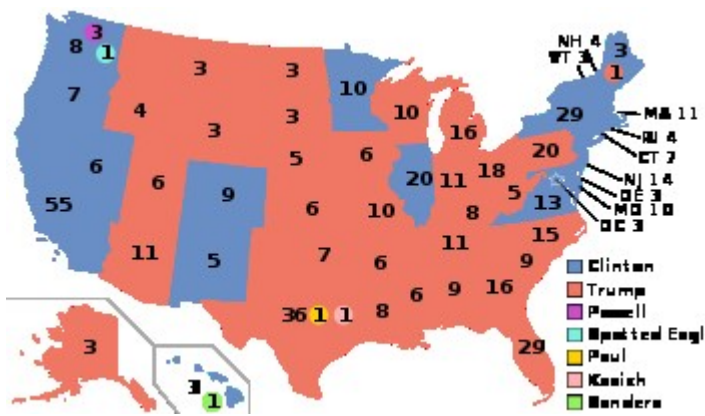
It's a striking figure, but it exaggerates the swings it's trying to depict. The picograms are scaled to the numbers in the little houses (which represent the number of homes built in the given months), but in both dimensions. And of course houses are three-dimensional objects, so that even though the picture doesn't depict the third dimension, our unconscious mind knows that these little domiciles have volume. So the Jan. 2006 house (2,273) is more than five times wider and higher than the April 2009 house (478). But five times in three dimensions:  $5 \times 5 \times 5 = 125$ . The Jan. 2006 house is over 125 times larger than the April 2009 house; that's why it looks like we have a mansion next to a shed. There were swings in housing construction over the years, but they weren't as large as this graphic makes them seem.

One ubiquitous picture that's easy to misinterpret, not because anybody broke Neurath's rules, but simply because of how things happen to be in the world, is the map of the United States. What makes it tricky is that the individual states' sizes are not proportional to their populations. This has the effect of exaggerating certain phenomena. Consider the final results of the 2016 presidential election, pictured, as they normally are, with states that went for the Republican candidate in red and those that went for the Democrat in blue. This is what you get<sup>63</sup>:

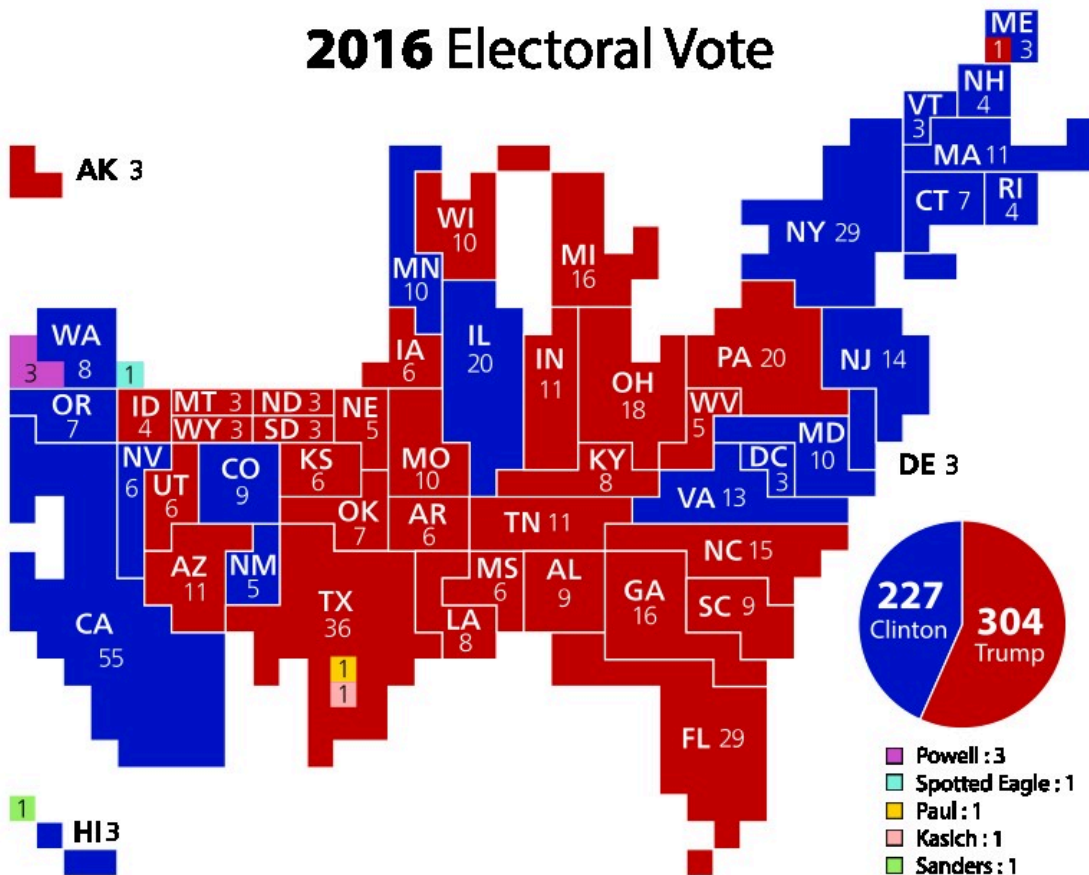
62. See here: [http://www.slate.com/blogs/moneybox/2011/12/23/america\\_s\\_housing\\_shortage.html](http://www.slate.com/blogs/moneybox/2011/12/23/america_s_housing_shortage.html)

63. Source of image: [https://en.wikipedia.org/wiki/Electoral\\_College\\_\(United\\_States\)](https://en.wikipedia.org/wiki/Electoral_College_(United_States))





Look at all that red! Clinton apparently got trounced. Except she didn't: she won the popular vote by more than three million. It looks like there are a lot more Trump votes because he won a lot of states that are very large but contain very few voters. Those Great Plains states are huge, but hardly anybody lives up there. If you were to adjust the map, making the states' sizes proportional to their populations, you'd end up with something like this<sup>64</sup>:



And this is only a partial correction: this sizes the states by electors in the Electoral College; that still exaggerates the sizes of some of those less-populated states. A true adjustment would have to show more blue than red, since Clinton won more votes overall.

I'll finish with an example stolen directly from the inspiration for this section—Darrell Huff's *How to Lie with Statistics*.<sup>65</sup>

64. Ibid.

65. p. 103.



It is a map of the United States made to raise alarm over the amount of spending being done by the federal government (it was produced over half a century ago; some things never change). Here it is:

## The Darkening Shadow

Federal Spending = Incomes of All People in Shaded States



That makes it look like federal spending is the equivalent of half the country's incomes! But Huff produced his own map ("Eastern style"), shading different states, same total population:



Not nearly so alarming.

People try to fool you in so many different ways. The only defense is a little logic, and a whole lot of skepticism. Be vigilant!