



# FOR ALL X

MIDWEST COLLABORATION

This is version 0.2 of An Open Introduction to Logic. It is current as of July 15, 2019.

© 2005–2019 by Cathal Woods, P.D. Magnus, J. Robert Loftis, and Adam Edwards. Some rights reserved.

Licensed under a Creative Commons license.

(Attribution-NonCommercial-ShareAlike 4.0 International ) <https://creativecommons.org/licenses/by-nc-sa/4.0/>



This book incorporates material from *An Introduction to Reasoning* by Cathal Woods, available at [sites.google.com/site/anintroductiontoreasoning/](https://sites.google.com/site/anintroductiontoreasoning/) and *For All X* by P.D. Magnus (version 1.27 [090604]), available at [www.fecundity.com/logic](http://www.fecundity.com/logic).

*Introduction to Reasoning* © 2007–2014 by Cathal Woods. Some rights reserved.

Licensed under a Creative Commons license: Attribution-NonCommercial-ShareAlike 3.0

Unported. <http://creativecommons.org/licenses/by-nc-sa/3.0/>

*For All X* © 2005–2010 by P.D. Magnus. Some rights reserved.

Licensed under a Creative Commons license: Attribution ShareAlike <http://creativecommons.org/licenses/by-sa/3.0/>

Adam Edwards compiled this edition and wrote original material for it. He takes full responsibility for any mistakes remaining in this version of the text.

Typesetting was carried out entirely in  $\text{\LaTeX}$ . The style for typesetting proofs is based on `fitch.sty` (v0.4) by Peter Selinger, University of Ottawa.

“When you come to any passage you don’t understand, *read it again*: if you *still* don’t understand it, *read it again*: if you fail, even after *three* readings, very likely your brain is getting a little tired. In that case, put the book away, and take to other occupations, and next day, when you come to it fresh, you will very likely find that it is *quite* easy.”

– Charles Dodgson (Lewis Carroll) *Symbolic Logic* (?)

“Few persons care to study logic, because everybody conceives himself to be proficient enough in the art of reasoning already. But I observe that this satisfaction is limited to one’s own ratiocination and does not extend to that of other men. We come to the full possession of our power of drawing inferences the last of all our faculties, for it is not so much a natural gift as a long and difficult art.”

– Charles Sanders Peirce “The Fixation of Belief”, in *Popular Science Monthly*, Vol. 12 (November 1877)”



# Contents

ABOUT THIS BOOK . . . . .	9
ACKNOWLEDGMENTS . . . . .	11
DEDICATION . . . . .	15
<b>I Basic Concepts</b>	<b>21</b>
1. WHAT IS LOGIC? . . . . .	23
Introduction . . . . .	23
Statement, Argument, Premise, Conclusion . . . . .	27
Arguments and Nonarguments . . . . .	34
2. THE BASICS OF EVALUATING ARGUMENT . . . . .	39
Two Ways an Argument Can Go Wrong . . . . .	39
Validity and Soundness . . . . .	40
Strong, Cogent, Deductive, Inductive . . . . .	45
3. WHAT IS FORMAL LOGIC? . . . . .	53
Formal as in Concerned with the Form of Things . . . . .	53
Formal as in Strictly Following Rules . . . . .	55
More Logical Notions for Formal Logic . . . . .	58
<b>II Critical Thinking</b>	<b>63</b>
4. WHAT IS CRITICAL THINKING? . . . . .	65
5. INFORMAL FALLACIES . . . . .	67
What are informal fallacies? . . . . .	67
Fallacy of Composition . . . . .	68
Fallacy of Division . . . . .	70
Begging the question . . . . .	70
False dichotomy . . . . .	72

Equivocation . . . . .	73
Slippery slope fallacies . . . . .	74
Fallacies of relevance . . . . .	78
Tu quoque . . . . .	82
6. COGNITIVE BIASES . . . . .	<b>85</b>
Introduction . . . . .	85
Confirmation Bias . . . . .	86
Availability Heuristic . . . . .	88
Anchoring and Ordering Effects . . . . .	88
Survivorship Bias . . . . .	89
Fundamental Attribution Error . . . . .	91
Emotive Biases . . . . .	91
Zeigarnik Effect . . . . .	92
<b>III Categorical Logic</b>	<b>95</b>
7. CATEGORICAL STATEMENTS . . . . .	<b>97</b>
Quantified Categorical Statements . . . . .	97
Quantity, Quality, Distribution, and Venn Diagrams . . . . .	99
Transforming English into Logically Structured English . . . . .	103
8. CATEGORICAL SYLLOGISMS . . . . .	<b>111</b>
Standard Form, Mood, and Figure . . . . .	111
Testing Validity . . . . .	114
9. EXISTENTIAL IMPORT . . . . .	<b>123</b>
Existential Import and Conditionally Valid Forms . . . . .	123
Rules and Fallacies . . . . .	127
Validity and the Counterexample Method . . . . .	134
Ordinary Language Arguments . . . . .	137
Enthymemes . . . . .	142
Sorites Categorical Arguments . . . . .	147
<b>IV Propositional Logic</b>	<b>157</b>
10. PROPOSITIONAL LOGIC . . . . .	<b>159</b>
Sentence Letters . . . . .	159
Sentential Connectives . . . . .	161
Negation . . . . .	162
Conjunction . . . . .	164

Disjunction . . . . .	166
Conditional . . . . .	167
Biconditional . . . . .	169
Combining connectives . . . . .	169
Unless . . . . .	170
Only . . . . .	171
Combining negation with conjunction and disjunction . . . . .	174
Recursive Syntax for SL . . . . .	175
Notational conventions . . . . .	177
<b>11. TRUTH TABLES . . . . .</b>	<b>181</b>
Basic Concepts . . . . .	181
Complete Truth Tables . . . . .	182
 <b>V Inductive Logic and Scientific Reasoning . . . . .</b>	 <b>187</b>
<b>12. WHAT ARE INDUCTION AND SCIENTIFIC REASONING? . . . . .</b>	<b>189</b>
Introduction . . . . .	189
Early Scientific Method . . . . .	191
The Hypothetico-Deductive Model . . . . .	192
The Demarcation Criterion . . . . .	194
When should we reject a hypothesis? . . . . .	196
Causation . . . . .	196
<b>13. INDUCTIVE LOGIC . . . . .</b>	<b>199</b>
What is probability? . . . . .	199
The Axioms of Probability . . . . .	200
<b>14. MILL'S METHODS . . . . .</b>	<b>203</b>
Direct method of agreement . . . . .	203
Method of Difference . . . . .	204
Joint method of agreement and difference . . . . .	204
Method of Residue . . . . .	205
Method of Concomitant Variation . . . . .	205
 <b>VI Appendices . . . . .</b>	 <b>207</b>
A. GLOSSARY . . . . .	211
B. OTHER SYMBOLIC NOTATION . . . . .	221
C. QUICK REFERENCE . . . . .	225

Characteristic Truth Tables . . . . .	225
Symbolization . . . . .	225
Using identity to symbolize quantities . . . . .	226



## *About this Book*

This book was created by combining two previous books on logic and critical thinking, both made available under a Creative Commons license, and then adding some material so that the coverage matched that of commonly used logic textbooks.

P.D. Magnus' *For All X* (?) formed the basis of Part ??: Formal Logic. I began using *For All X* in my own logic classes in 2009, but I quickly realized I needed to make changes to make it appropriate for the community college students I was teaching. In 2010 I began developing *For All X : The Lorain County Remix* and using it in my classes. The main change I made was to separate the discussions of sentential and quantificational logic and to add exercises. It is this remixed version that became the basis for Part ??: Formal Logic complete version of this text.

Similarly, Part ??: Critical Thinking and Part V: Inductive and Scientific Reasoning grew out of Cathal Woods' *Introduction to Reasoning*. In the Spring of 2011, I began to use an early version of this text <sup>(1)</sup> in my critical thinking courses. I kept up with the updates and changes to the text until the release of <sup>2</sup>, all the while gradually merging the material with the work in *For All X*. After that point, my version forks from Woods's.

On May 20, 2016, I posted the combined textbook to Github and all subsequent changes have been tracked there: <https://github.com/rob-helpey-chalk/openintroduction>

J. Robert Loftis  
Elyria, Ohio, USA



## *Acknowledgments*

Thank you to all of the mentors and students that have supported this project and have helped it become a reality. Special thanks to Jonathan Livengood for the inspiration to write a free logic textbook and to Ashli Anda for her inexhaustable support.

Adam Edwards  
*Champaign, Illinois*

Thanks first of all go to the authors of the textbooks here stitched together: P.D. Magnus for *For All X* and Cathal Woods for *Introduction to Reasoning*. My thanks go to them for writing the excellent textbooks that have been incorporated into this one, for making those publicly available under Creative Commons licenses, and for giving their blessing to this derivative work.

In general, this book would not be possible without a culture of sharing knowledge. The book was typeset using  $\text{\LaTeX}$  developed by Leslie Lamport. Lamport was building on  $\text{\TeX}$  by Donald Knuth. Peter Selinger built on what Lamport made by developing the Fitch typesetting format that the proofs were laid out in. Diagrams were made in TikZ by Till Tantau. All of these are coding systems are not only freely available online, they have extensive user support communities. Add-on packages are designed, manuals are written, questions are answered in discussion forums, all by people who are donating their time and expertise.

The culture of sharing isn't just responsible for the typesetting of this book; it was essential to the content. Essential background information comes from the free online *Stanford Encyclopedia of Philosophy*. Primary sources from the history of logic came from *Project Gutenberg*. Logicians, too, can and should create free knowledge.

Many early adopters of this text provided invaluable feedback, including Jeremy Dolan, Terry Winant, Benjamin Lennertz, Ben Sheredos, and Michael Hartsock. Lennertz, in particular, provided useful edits. Helpful comments were also made by Ben Cordry, John Emerson, Andrew Mills, Nathan Smith, Vera Tobin, Cathal Woods, and many more that I have forgot to mention, but whose emails are probably sitting on my computer somewhere.

I would also like to thank Lorain County Community College for providing the sabbatical leave that allowed me to write the sections of this book on Aristotelian logic. Special thanks goes to all the students at LCCC who had to suffer through earlier versions of this work and provided much helpful feedback. Most importantly, I would like to thank Molly, Caroline and Joey for their incredible love and support.

J. Robert Loftis  
*Elyria, Ohio, USA*

Intellectual debts too great to articulate are owed to scholars too many to enumerate. At different points in the work, readers might detect the influence of various works of Aristotle, Toulmin (especially <sup>3</sup>), Fisher and Scriven (?), Walton (especially <sup>4</sup>), Epstein (?), Johnson-Laird (especially (?)), Scriven (?), Giere (?) and the works of the Amsterdam school of pragma-dialectics (?).

Thanks are due to Virginia Wesleyan College for providing me with Summer Faculty Development funding in 2008 and 2010 and a Batten professorship in 2011. These funds, along with some undergraduate research funds (also provided by VWC), allowed me to hire students Gaby Alexander (2008), Ksera Dyette (2009), Mark Jones (2008), Andrea Medrano (2011), Lauren Perry (2009), and Alan Robertson (2010). My thanks to all of them for their hard work and enthusiasm.

For feedback on the text, thanks are due to James Robert (Rob) Loftis (Lorain County Community College) and Bill Roche (Texas Christian University). Answers (to exercises) marked with “(JRL)” are by James Robert Loftis.

Particular thanks are due to my (once) Ohio State colleague Bill Roche. The book began as a collection of lecture notes, combining work by myself and Bill.

Cathal Woods

*Norfolk, Virginia, USA*

(Taken from *Introduction to Reasoning* (?))

The author would like to thank the people who made this project possible. Notable among these are Cristyn Magnus, who read many early drafts; Aaron Schiller, who was an early adopter and provided considerable, helpful feedback; and Bin Kang, Craig Erb, Nathan Carter, Wes McMichael, and the students of Introduction to Logic, who detected various errors in previous versions of the book.

P.D. Magnus

*Albany, New York, USA*

(Taken from *For All X* (?))



## *Dedication*

*This book is dedicated to Arlo, that you may not be deceived.*





## List of Figures

1.1	The boardgame Clue. . . . .	23
1.2	The ob/ob mouse (left), a laboratory mouse which has been genetically engineered to be obese, and an ordinary mouse (right). Photo from ?. . . . .	26
1.3	A statement in different contexts, or no context. . . . .	28
2.1	Neither argument is valid, but one is strong and one is weak . . . . .	49
3.1	Two examples of valid arguments. . . . .	53
6.1	A diagram of the locations where damage was recorded on bombers during WWII. Source: <a href="https://commons.wikimedia.org/wiki/File:Survivorship-bias.png">https://commons.wikimedia.org/wiki/File:Survivorship-bias.png</a> . . . . .	90
7.1	Parts of a quantified categorical statement. . . . .	98
7.2	Euler diagrams from Scottish philosopher William Hamilton's <i>Lectures on Logic</i> . . . . .	101
7.3	Venn's original diagram for an mood-A statement (Venn (?)). Screenshot from Google Books by J. Robert Loftis. . . . .	101
7.4	A basic Venn diagram. . . . .	102
7.5	A Venn diagram for a mood-A sentence. . . . .	102
7.6	A Venn diagram for a mood-I sentence. . . . .	102
7.7	Venn diagrams for all four categorical moods. . . . .	103
8.1	The four Venn diagrams for mood A sentences. . . . .	116
8.2	The two Venn diagrams for mood E sentences. . . . .	116
8.3	It is ambiguous where the X should go on this Venn diagram. . . . .	116
8.4	We place the X on the border between these regions to indicate that it <i>could</i> be in either. . . . .	116
8.5	The X is forced into region 3 because we know that nothing can be in region 4. . . . .	117
8.6	The X is forced into region 3 because we know that nothing can be in region 4. . . . .	117
8.7	Basic Venn diagrams for all particular categorical sentences. Note that syllogisms with universal sentences may shade some regions and force the X off the border and completely into one of the regions. . . . .	118
8.8	"All M are P." . . . . .	118
8.9	"All M are P." and "All S are M." . . . . .	118
8.10	"All P are M." . . . . .	119
8.11	"All P are M." and "All S are M." . . . . .	119
8.12	"All S are M." . . . . .	120
8.13	"Some M are P." and "All S are M." . . . . .	120

8.14	“All $M$ are $S$ ” . . . . .	120
8.15	“Some $M$ are $P$ ” and “All $M$ are $S$ ” . . . . .	120
8.16	The oldest surviving version of the “Barbara, Celarent...” poem, from William of Sherwood. The manuscript is held at the Bibliothèque Nationale de France, ms. Lat. 16617, <a href="http://gallica.bnf.fr/ark:/12148/btv1b9066740r">http://gallica.bnf.fr/ark:/12148/btv1b9066740r</a> . . . . .	121
10.1	This is Cupcake. The fact that she is a dog is a <i>sufficient</i> condition for her to be a mammal. She also likes socks. . . . .	173
10.2	The antecedent of a material conditional is a sufficient condition for the consequent, while the consequent is a necessary condition for the antecedent. . . . .	173

## *List of Tables*

1.2	Premise and Conclusion Indicators. . . . .	31
7.1	The four moods of categorical statements . . . . .	100
7.2	Quantity, quality, and distribution. . . . .	100
8.1	The sixty-four moods of the Aristotelian syllogism . . . . .	112
8.2	The four figures of the Aristotelian syllogism . . . . .	113
8.3	The 15 unconditionally valid syllogisms. . . . .	121
9.2	All 24 Valid Syllogisms . . . . .	125
9.3	Moods and distribution . . . . .	152
10.1	The Sentential Connectives. . . . .	162
10.2	The basic elements of SL . . . . .	176
11.2	The characteristic truth tables for the connectives of SL. . . . .	183
14.1	The Method of Agreement . . . . .	204
14.2	The Method of Difference . . . . .	204
14.3	The Joint Method . . . . .	205
14.4	The Method of Residue . . . . .	205
14.5	The Method of Concomitant Variation . . . . .	206



## **Part I**

# **Basic Concepts**



# 1

## *What Is Logic?*

### *Introduction*

Logic is a part of the study of human reasoning—the ability we have to think abstractly, solve problems, explain the things that we know, and infer new knowledge on the basis of evidence. Traditionally, logic has focused on the last of these items, the ability to make inferences on the basis of evidence. This is an activity you engage in every day. Consider, for example, the game of Clue. (For those of you who have never played, Clue is a murder mystery game where players have to decide who committed the murder, what weapon they used, and where they were.) A player in the game might decide that the murder weapon was the candlestick by ruling out the other weapons in the game: the knife, the revolver, the rope, the lead pipe, and the wrench. This evidence lets the player know something they did not know previously, namely, the identity of the murderer.

In logic, we use the word “argument” to refer to the attempt to show that certain evidence supports a conclusion. This is very different from the sort of argument you might have when you are mad at someone, which could involve a lot of yelling. We are going to use the word “argument” a lot in this book, so you need to get used to thinking of it as a name for an abstract and rational process, and not a word that describes what happens when people disagree.

A logical argument is structured to give someone a reason to believe some conclusion. Here is the argument about a game of Clue written out in a way that shows its structure.

- (A1) In a game of Clue, the possible murder weapons are the knife, the candlestick, the revolver, the rope, the lead pipe, and the wrench.
- (A2) The murder weapon was not the knife.
- (A3) The murder weapon was also not the revolver, the rope, the lead pipe, or the wrench.
- (A4)  $\therefore$  Therefore, the murder weapon was the candlestick.

In the argument above, statements A1–A3 are the evidence. We call

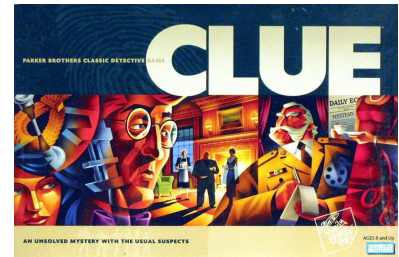


Figure 1.1: The boardgame Clue.

these the **PREMISES**. The word “therefore” indicates that the final statement, marked with a C, is the **CONCLUSION** of the argument. If you believe the premises, then the argument provides you with a reason to believe the conclusion. You might use reasoning like this purely in your own head, without talking with anyone else. You might wonder what the murder weapon is, and then mentally rule out each item, leaving only the candlestick. On the other hand, you might use reasoning like this while talking to someone else, to convince them that the murder weapon is the candlestick. (Perhaps you are playing as a team.) Either way the structure of the reasoning is the same.

We can define **LOGIC** then more precisely as the part of the study of the structure of good inferences. In more casual situations, we will follow ordinary practice and use the word “logic” to either refer to the business of studying inferences and arguments in general or the thing being studied, that is, inference itself. While logic focuses on inferential reasoning, other disciplines, like decision theory and cognitive science, deal with other aspects of human reasoning, like abstract thinking and problem solving more generally. Logic, as the study of inference, has been pursued for thousands of years by people from civilizations all over the globe. The initial motivation for studying logic is often practical. Given that we use arguments and make inferences all the time, it only makes sense that we would want to learn to do these things better. Once people begin to study logic, however, they often realize that it is a fascinating topic in its own right. Thus, the study of logic quickly moves from being a practical business to a theoretical endeavor people pursue for its own sake.

In order to study reasoning, we have to apply our ability to reason to our reason itself. This reasoning about reasoning is called **METAREASONING**. It is part of a more general set of processes called **METACOGNITION**, which is just any kind of thinking about thinking. When we are pursuing logic as a practical discipline, one important part of metacognition will be awareness of your own thinking, especially its weakness and biases, as it is occurring. More theoretical metacognition will be about attempting to understand the structure of thought itself.

Whether we are pursuing logical for practical or theoretical reasons, our focus is on argument. The key to studying argument is to set aside the subject being argued about and to focus on the *way* it is argued *for*. The section opened with an example that was about a game of Clue. However, the kind of reasoning used in that example was just the process of elimination. Process of elimination can be applied to any subject. Suppose a group of friends is deciding which restaurant to eat at, and there are six restaurants in town. If you could rule out five of the possibilities, you would use an argument just like the one above to decide where to eat. Because logic sets aside what an argument is about, and just looks at how it works rationally, logic is said to have **CONTENT NEUTRALITY**. If we say an argument is good, then the

Throughout the book, we will also use the triangle of dots symbol “.” to indicate that a statement is a conclusion.



same kind of argument applied to a different topic will also be good. If we say an argument is good for solving murders, we will also say that the same kind of argument is good for deciding where to eat, what kind of disease is destroying your crops, or who to vote for.

When logic is studied for theoretical reasons, it typically is pursued as **FORMAL LOGIC**. In formal logic we get content neutrality by replacing parts of the argument we are studying with abstract symbols. For instance, we could turn the argument above into a formal argument like this:

- (B1) There are six possibilities: A, B, C, D, E, and F.
- (B2) A is false.
- (B3) B, D, E, and F are also false.
- (B4)  $\therefore$  The correct answer is C.

Here we have replaced the concrete possibilities in the first argument with abstract letters that could stand for anything. We have also replaced the English word “therefore” with the symbol “ $\therefore$ ,” which means therefore. This lets us see the formal structure of the argument, which is why it works in any domain you can think of. In fact, we can think of formal logic as the method for studying argument that uses abstract notation to identify the formal structure of argument. Formal logic is closely allied with mathematics, and studying formal logic often has the sort of puzzle-solving character one associates with mathematics. You will see this when we get to Part ??, which covers formal logic.

When logic is studied for practical reasons, it is typically called critical thinking. We will define **CRITICAL THINKING** narrowly as the use of metareasoning to improve our reasoning in practical situations. Sometimes we will use the term “critical thinking” more broadly to refer to the results of this effort at self-improvement. You are “thinking critically” when you reason in a way that has been sharpened by reflection and metareasoning. A **CRITICAL THINKER** someone who has both sharpened their reasoning abilities using metareasoning and deploys those sharpened abilities in real world situations.

Critical thinking is generally pursued as **INFORMAL LOGIC**, rather than formal logic. This means that we will keep arguments in ordinary language and draw extensively on your knowledge of the world to evaluate them. In contrast to the clarity and rigor of formal logic, informal logic is suffused with ambiguity and vagueness. There are problems with multiple correct answers, or where reasonable people can disagree with what the correct answer is. This is because you will be dealing with reasoning in the real world, which is messy.

You can think of the difference between formal logic and informal logic as the difference between a laboratory science and a field science. If you are studying, say, mice, you could discover things about them by running experiments in a lab, or you can go out into the field where mice live and

observe them in their natural habitat. Informal logic is the field science for arguments: you go out and study arguments in their natural habitats, like newspapers, courtrooms, and scientific journal articles. Like studying mice scurrying around a meadow, the process takes patience, and often doesn't yield clear answers but it lets you see how things work in the real world. Formal logic takes arguments out of their natural habitat and performs experiments on them to see what they are capable of. The arguments here are like lab mice. They are pumped full of chemicals and asked to perform strange tasks, as it were. They live lives very different than their wild cousins. Some of the arguments will wind up looking like the "ob/ob mouse", a genetically engineered obese mouse scientists use to study type II diabetes (See Figure 1.2). These arguments will be huge, awkward, and completely unable to survive in the wild. But they will tell us a lot about the limits of logic as a process.



Figure 1.2: The ob/ob mouse (left), a laboratory mouse which has been genetically engineered to be obese, and an ordinary mouse (right). Photo from ?.

Our main goal in studying arguments is to separate the good ones from the bad ones. The argument about Clue we saw earlier is a good one, based on the process of elimination. It is good because it leads to truth. If I've got all the premises right, the conclusion will also be right. The textbook *Logic: Techniques of Formal Reasoning* (?) had a nice way of capturing the meaning of logic: "logic is the study of virtue in argument." This textbook will accept this definition, with the caveat that an argument is virtuous if it helps us get to the truth.

Logic is different from **RHETORIC**, which is the study of effective persuasion. Rhetoric does not look at virtue in argument. It only looks at the power of arguments, regardless of whether they lead to truth. An advertise-

ment might convince you to buy a new truck by having a gravelly voiced announcer tell you it is “ram tough” and showing you a picture of the truck on top of a mountain, where it no doubt actually had to be airlifted. This sort of persuasion is often more effective at getting people to believe things than logical argument, but it has nothing to do with whether the truck is really the right thing to buy. In this textbook we will only be interested in rhetoric to the extent that we need to learn to defend ourselves against the misleading rhetoric of others. This will not, however, be anything close to a full treatment of the study of rhetoric.

### *Statement, Argument, Premise, Conclusion*

So far we have defined logic as the study of argument and outlined its relationship to related fields. To go any further, we are going to need a more precise definition of what exactly an argument is. We have said that an argument is not simply two people disagreeing; it is an attempt to prove something using evidence. More specifically, an argument is composed of statements. In logic, we define a **STATEMENT** as a unit of language that can be true or false. That means that statements are **TRUTH EVALUABLE**. All of the items below are statements.

1. *Tyrannosaurus rex* went extinct 65 million years ago.
2. *Tyrannosaurus rex* went extinct last week.
3. On this exact spot, 100 million years ago, a *T. rex* laid a clutch of eggs.
4. Abraham Lincoln is the king of Mars.
5. Murder is wrong.
6. Abortion is murder.
7. Abortion is a woman’s right.
8. Lady Gaga is pretty.
9. Murder is the unjustified killing of a person.
10. The slithy toves did gyre and gimble in the wabe.
11. The murder of logician Richard Montague was never solved.

Because a statement is something that can be true *or* false, statements include truths like 1 and falsehoods like 2. A statement can also be something that must either be true or false, but we don’t know which, like 3. A statement can be something that is completely silly, like 4. Statements in logic include statements about morality, like 5, and things that in other contexts might be called “opinions,” like 6 and 7. People disagree strongly

about whether 6 or 7 are true, but it is definitely possible for one of them to be true. The same is true about 8, although it is a less important issue than 6 and 7. A statement in logic can also simply give a definition, like 9. This sort of statement announces that we plan to use words a certain way, which is different from statements that describe the world, like 1, or statements about morality, like 6. Statements can include nonsense words like 10, because we don't really need to know what the statement is about to see that it is the sort of thing that can be true or false. All of this relates back to the content neutrality of logic. The statements we study can be about dinosaurs, abortion, Lady Gaga, and even the history of logic itself, as in statement 11, which is true.

We are treating statements primarily as units of language or strings of symbols, and most of the time the statements you will be working with will just be words printed on a page. However, it is important to remember that statements are also what philosophers call “speech acts.” They are actions people take when they speak (or write). If someone makes a statement they are typically telling other people that they believe the statement to be true, and will back it up with evidence if asked to. When people make statements, they always do it in a context—they make statements at a place and a time with an audience. Often the context statements are made in will be important for us, so when we give examples, statements, or arguments we will sometimes include a description of the context. When we do that, we will give the context in *italics*. See Figure 1.3 for examples. For the most part, the context for a statement or argument will be important in the chapters on critical thinking, when we are pursuing the study of logic for practical reasons. In the chapters on formal logic, context is less important, and we will be more likely to skip it.

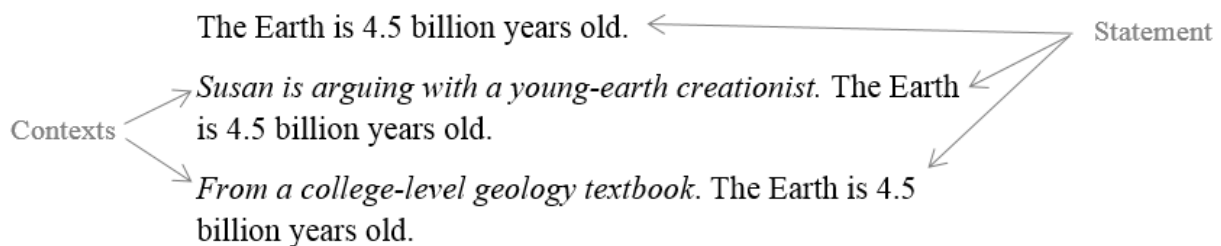


Figure 1.3: A statement in different contexts, or no context.

“Statements” in this text will *not* include questions, commands, exclamations, or sentence fragments. Someone who asks a *question* like “Does the grass need to be mowed?” is typically not claiming that anything is true or false. *Questions* do not count as statements, but *answers* usually will. “What is this course about?” is not a statement. An answer to that question such as “No one knows what this course is about,” is a statement.

For the same reason *commands* do not count as statements for us. If someone bellows “Mow the grass, now!” they are not saying whether the grass has been mowed or not. You might infer that they believe the lawn has not been mowed, but then again maybe they think the lawn is fine and just want to see you exercise.

An exclamation like “Ouch!” is also neither true nor false. On its own, it is not a statement. We will treat “Ouch, I hurt my toe!” as meaning the same thing as “I hurt my toe.” The “ouch” does not add anything that could be true or false.

Finally, a lot of possible strings of words will fail to qualify as statements simply because they don’t form a complete sentence. In your composition classes, these were probably referred to as sentence fragments. This includes strings of words that are parts of sentences, such as noun phrases like “The tall man with the hat” and verb phrases, like “ran down the hall.” Phrases like these are missing something they need to make a claim about the world. The class of sentence fragments also includes completely random combinations of words, like “The up if blender route,” which don’t even have the form of a statement about the world.

Other logic textbooks describe the components of argument as “propositions,” or “assertions,” and we will use these terms sometimes as well. There is actually a great deal of disagreement about what the differences between all of these things are and which term is best used to describe parts of arguments. However, none of that makes a difference for this textbook. Some textbooks will also use the term “sentence” here. We will not use the word “sentence” to mean the same thing as “statement.” Instead, we will use “sentence” the way it is used in ordinary grammar, to refer generally to statements, questions, and commands.

Sometimes the outward form of a speech act does not match how it is actually being used. A rhetorical question, for instance, has the outward form of a question, but is really a statement or a command. If someone says “don’t you think the lawn needs to be mowed?” they may actually mean a statement like “the lawn needs to be mowed” or a command like “mow the lawn, now.” Similarly one might disguise a command as a statement. “You will respect my authority” is either true or false—either you will or you will not. But the speaker may intend this as an order—“Respect me!”—rather than a prediction of how you will behave.

When we study argument, we need to express things as statements, because arguments are composed of statements. Thus if we encounter a rhetorical question while examining an argument, we need to convert it into a statement. “Don’t you think the lawn needs to be mowed” will become “the lawn needs to be mowed.” Similarly, commands will become should statements. “Mow the lawn, now!” will need to be transformed into “You should mow the lawn.”

The latter kind of change will be important in critical thinking, because

critical thinking often studies arguments whose goal is to get an audience to do something. These are called **PRACTICAL ARGUMENTS**. Most advertising and political speech consists of practical arguments, and these are crucial topics for critical thinking.

Once we have a collection of statements, we can use them to build arguments. An **ARGUMENT** is a connected series of statements designed to convince an audience of another statement. Here an audience might be a literal audience sitting in front of you at some public speaking engagement. Or it might be the readers of a book or article. The audience might even be yourself as you reason your way through a problem. Let's start with an example of an argument given to an external audience. This passage is from an essay by Peter Singer called "Famine, Affluence, and Morality" in which he tries to convince people in rich nations that they need to do more to help people in poor nations who are experiencing famine.

*A contemporary philosopher writing in an academic journal.* If it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do so. Famine is something bad, and it can be prevented without sacrificing anything of comparable moral importance. So, we ought to prevent famine. ?

Singer wants his readers to work to prevent famine. This is represented by the last statement of the passage, "we ought to prevent famine," which is called the conclusion of the passage. The **CONCLUSION** of an argument is the statement that the argument is trying to convince the audience of. The statements that do the convincing are called the **PREMISES**. In this case, the argument has three premises: (1) "If it is in our power to prevent something bad from happening, without thereby sacrificing anything of comparable moral importance, we ought, morally, to do so"; (2) "Famine is something bad"; and (3) "it can be prevented without sacrificing anything of comparable moral importance."

Now let's look at an example of internal reasoning.

*Jack arrives at the track, in bad weather.* There is no one here. I guess the race is not happening.

In the passage above, the words in *italics* explain the context for the reasoning, and the words in regular type represent what Jack is actually thinking to himself. (We will talk more about his way of representing reasoning in section ??, below.) This passage again has a premise and a conclusion. The premise is that no one is at the track, and the conclusion is that the race was canceled. The context gives another reason why Jack might believe the race has been canceled, the weather is bad. You could view this as another premise—it is very likely a reason Jack has come to believe that the race is canceled. In general, when you are looking at people's internal reasoning, it is often hard to determine what is actually

working as a premise and what is just working in the background of their unconscious.

When people give arguments to each other, they typically use words like “therefore” and “because.” These are meant to signal to the audience that what is coming is either a premise or a conclusion in an argument. Words and phrases like “because” signal that a premise is coming, so we call these **PREMISE INDICATORS**. Similarly, words and phrases like “therefore” signal a conclusion and are called **CONCLUSION INDICATORS**. The argument from Peter Singer (on page 30) uses the conclusion indicator word, “so.” Table 1.2 is an incomplete list of indicator words and phrases in English.

<b>Premise Indicators:</b>	because, as, for, since, given that, for the reason that
<b>Conclusion Indicators:</b>	therefore, thus, hence, so, consequently, it follows that, in conclusion, as a result, then, must, accordingly, this implies that, this entails that, we may infer that

Table 1.2: Premise and Conclusion Indicators.

The two passages we have looked at in this section so far have been simply presented as quotations. But often it is extremely useful to rewrite arguments in a way that makes their logical structure clear. One way to do this is to use something called “canonical form.” An argument written in **CANONICAL FORM** has each premise numbered and written on a separate line. Indicator words and other unnecessary material should be removed from the premises. Although you can shorten the premises and conclusion, you need to be sure to keep them all complete sentences with the same meaning, so that they can be true or false. The argument from Peter Singer, above, looks like this in canonical form:

- (C1) If we can stop something bad from happening, without sacrificing anything of comparable moral importance, we ought to do so.
- (C2) Famine is something bad.
- (C3) Famine can be prevented without sacrificing anything of comparable moral importance.
- (C4) ∴ We ought to prevent famine.

Each statement has been written on its own line and given a number. The statements have been paraphrased slightly, for brevity, and the indicator word “so” has been removed. Also notice that the “it” in the third premise has been replaced by the word “famine,” so that statements reads naturally on its own.

Similarly, we can rewrite the argument Jack gives at the racetrack, on page 30, like this:

(D1) There is no one at the race track.

(D2)  $\therefore$  The race is not happening.

Notice that we did not include anything from the part of the passage in italics. The italics represent the context, not the argument itself. Also, notice that the “I guess” has been removed. When we write things out in canonical form, we write the content of the statements, ignore information about the speaker’s mental state, like “I believe” or “I guess.”

One of the first things you have to learn to do in logic is to identify arguments and rewrite them in canonical form. This is a foundational skill for everything else we will be doing in this text, so we are going to run through a few examples now, and there will be more in the exercises. The passage below is paraphrased from the ancient Greek philosopher Aristotle.

*An ancient philosopher, writing for his students* Again, our observations of the stars make it evident that the earth is round. For quite a small change of position to south or north causes a manifest alteration in the stars which are overhead. (<sup>1</sup>, 298a2-10)

1

The first thing we need to do to put this argument in canonical form is to identify the conclusion. The indicator words are the best way to do this. The phrase “make it evident that” is a conclusion indicator phrase. He is saying that everything else is *evidence* for what follows. So we know that the conclusion is that the earth is round. “For” is a premise indicator word—it is sort of a weaker version of “because.” Thus the premise is that the stars in the sky change if you move north or south. In canonical form, Aristotle’s argument that the earth is round looks like this.

(E1) There are different stars overhead in the northern and southern parts of the earth.

(E2)  $\therefore$  The earth is spherical in shape.

That one is fairly simple, because it just has one premise. Here’s another example of an argument, this time from the book of Ecclesiastes in the Bible. The speaker in this part of the bible is generally referred to as The Preacher, or in Hebrew, Koheleth. In this verse, Koheleth uses both a premise indicator and a conclusion indicator to let you know he is giving reasons for enjoying life.

*The words of the Preacher, son of David, King of Jerusalem* There is something else meaningless that occurs on earth: the righteous who get what the wicked deserve, and the wicked who get what the righteous deserve. ...So I commend the enjoyment of life, because there is nothing better for a person under the sun than to eat and drink and be glad. (Ecclesiastes 8:14-15, New International Version)

Koheleth begins by pointing out that good things happen to bad people and bad things happen to good people. This is his first premise. (Most



Bible teachers provide some context here by pointing that that the ways of God are mysterious and this is an important theme in Ecclesiastes.) Then Koheleth gives his conclusion, that we should enjoy life, which he marks with the word “so.” Finally he gives an extra premise, marked with a “because,” that there is nothing better for a person than to eat and drink and be glad. In canonical form, the argument would look like this.

- (F1) Good things happen to bad people and bad things happen to good people.  
 (F2) There is nothing better for people than to eat, to drink and to enjoy life.  
 (F3)  $\therefore$  You should enjoy life.

Notice that in the original passages, Aristotle put the conclusion in the first sentence, while Koheleth put it in the middle of the passage, between two premises. In ordinary English, people can put the conclusion of their argument where ever they want. However, when we write the argument in canonical form, the conclusion goes last.

Unfortunately, indicator words aren’t a perfect guide to when people are giving an argument. Look at this passage from a newspaper:

*From the general news section of a national newspaper* The new budget underscores the consistent and paramount importance of tax cuts in the Bush philosophy. His first term cuts affected more money than any other initiative undertaken in his presidency, including the costs thus far of the war in Iraq. All told, including tax incentives for health care programs and the extension of other tax breaks that are likely to be taken up by Congress, the White House budget calls for nearly \$300 billion in tax cuts over the next five years, and \$1.5 trillion over the next 10 years. (?)

Although there are no indicator words, this is in fact an argument. The writer wants you to believe something about George Bush: tax cuts are his number one priority. The next two sentences in the paragraph give you reasons to believe this. You can write the argument in canonical form like this.

- (G1) Bush’s first term cuts affected more money than any other initiative undertaken in his presidency, including the costs thus far of the war in Iraq.  
 (G2) The White House budget calls for nearly \$300 billion in tax cuts over the next five years, and \$1.5 trillion over the next 10 years.  
 (G3)  $\therefore$  Tax cuts are of consistent and paramount importance of in the Bush philosophy.

The ultimate test of whether something is an argument is simply whether some of the statements provide reason to believe another one of the statements. If some statements support others, you are looking at an argument.

The speakers in these two cases use indicator phrases to let you know they are trying to give an argument.

A final bit of terminology for this section. An **INFERENCE** is the act of coming to believe a conclusion on the basis of some set of premises. When Jack in the example above saw that no one was at the track, and came to believe that the race was not on, he was making an inference. We also use the term inference to refer to the connection between the premises and the conclusion of an argument. If your mind moves from premises to conclusion, you make an inference, and the premises and the conclusion are said to be linked by an inference. In that way inferences are like argument glue: they hold the premises and conclusion together.

### *Arguments and Nonarguments*

We just saw that arguments are made of statements. However, there are lots of other things you can do with statements. Part of learning what an argument is involves learning what an argument is not, so in this section and the next we are going to look at some other things you can do with statements besides make arguments.

The list below of kinds of nonarguments is not meant to be exhaustive: there are all sorts of things you can do with statements that are not discussed. Nor are the items on this list meant to be exclusive. One passage may function as both, for instance, a narrative and a statement of belief. Right now we are looking at real world reasoning, so you should expect a lot of ambiguity and imperfection.

### *Simple Statements of Belief*

An argument is an attempt to persuade an audience to believe something, using reasons. Often, though, when people try to persuade others to believe something, they skip the reasons, and give a **SIMPLE STATEMENT OF BELIEF**. This is a kind of nonargumentative passage where the speaker simply asserts what they believe without giving reasons. Sometimes simple statements of belief are prefaced with the words “I believe,” and sometimes they are not. A simple statements of belief can be a profoundly inspiring way to change people’s hearts and minds. Consider this passage from Dr. Martin Luther King’s Nobel acceptance speech.

I believe that even amid today’s mortar bursts and whining bullets, there is still hope for a brighter tomorrow. I believe that wounded justice, lying prostrate on the blood-flowing streets of our nations, can be lifted from this dust of shame to reign supreme among the children of men. I have the audacity to believe that peoples everywhere can have three meals a day for their bodies, education and culture for their minds, and dignity, equality and freedom for their spirits. (?)

This actually is a part of a longer passage that consists almost entirely of statements that begin with some variation of “I believe.” It is incredibly powerful oration, because the audience, feeling the power of King’s beliefs, comes to share in those beliefs. The language King uses to describe how he believes is important, too. He says his belief in freedom and equality requires audacity, making the audience feel his courage and want to share in this courage by believing the same things.

These statements are moving, but they do not form an argument. None of these statements provide evidence for any of the other statements. In fact, they all say roughly the same thing, that good will triumph over evil. So the study of this kind of speech belongs to the discipline of rhetoric, not of logic.

### *Expository Passages*

Perhaps the most basic use of a statement is to convey information. Often if we have a lot of information to convey, we will sometimes organize our statements around a theme or a topic. Information organized in this fashion can often appear like an argument, because all of the statements in the passage relate back to some central statement. However, unless the other statements are given as reasons to believe the central statement, the passage you are looking at is not an argument. Consider this passage:

*From a college psychology textbook.* Eysenck advocated three major behavior techniques that have been used successfully to treat a variety of phobias. These techniques are modeling, flooding, and systematic desensitization. In **modeling** phobic people watch nonphobics cope successfully with dreaded objects or situations. In **flooding** clients are exposed to dreaded objects or situations for prolonged periods of time in order to extinguish their fear. In contrast to flooding, **systematic desensitization** involves gradual, client-controlled exposure to the anxiety eliciting object or situation. (Adapted from Ryckman <sup>2</sup>Ryckman2007)

2

We call this kind of passage an expository passage. In an **EXPOSITORY PASSAGE**, statements are organized around a central theme or topic statement. The topic statement might look like a conclusion, but the other statements are not meant to be evidence for the topic statement. Instead, they elaborate on the topic statement by providing more details or giving examples. In the passage above, the topic statement is “Eysenck advocated three major behavioral techniques . . .” The statements describing these techniques elaborate on the topic statement, but they are not evidence for it. Although the audience may not have known this fact about Eysenck before reading the passage, they will typically accept the truth of this statement instantly, based on the textbook’s authority. Subsequent statements in the passage merely provide detail.

Deciding whether a passage is an argument or an expository passage is complicated by the fact that sometimes people argue by example:

- Steve:** Kenyans are better distance runners than everyone else.
- Monica:** Oh come on, that sounds like an exaggeration of a stereotype that isn't even true.
- Steve:** What about Dennis Kimetto, the Kenyan who set the world record for running the marathon? And you know who the previous record holder was: Emmanuel Mutai, also Kenyan.

Here Steve has made a general statement about all Kenyans. Monica clearly doubts this claim, so Steve backs it up with some examples that seem to match his generalization. This isn't a very strong way to argue: moving from two examples to statement about all Kenyans is probably going to be a kind of bad argument known as a hasty generalization. (This mistake is covered in the complete version of this text in the chapter on induction Chapter ?? on induction.) The point here however, is that Steve is just offering it as an argument.

The key to telling the difference between expository passages and arguments by example is whether there is a conclusion that they audience needs to be convinced of. In the passage from the psychology textbook, "Eysenck advocated three major behavioral techniques" doesn't really work as a conclusion for an argument. The audience, students in an introductory psychology course, aren't likely to challenge this assertion, the way Monica challenges Steve's overgeneralizing claim.

Context is very important here, too. The Internet is a place where people argue in the ordinary sense of exchanging angry words and insults. In that context, people are likely to actually give some arguments in the logical sense of giving reasons to believe a conclusion.

### *Narratives*

Statements can also be organized into descriptions of events and actions, as in this snippet from book V of *Harry Potter*.

But she [Hermione] broke off; the morning post was arriving and, as usual, the *Daily Prophet* was soaring toward her in the beak of a screech owl, which landed perilously close to the sugar bowl and held out a leg. Hermione pushed a Knut into its leather pouch, took the newspaper, and scanned the front page critically as the owl took off again. (?)

We will use the term **NARRATIVE** loosely to refer to any passage that gives a sequence of events or actions. A narrative can be fictional or nonfictional. It can be told in regular temporal sequence or it can jump around, forcing the audience to try to reconstruct a temporal sequence. A narrative can describe a short sequence of actions, like Hermione taking a newspaper from an owl, or a grand sweep of events, like this passage about the rise and fall of an empire in the ancient near east:

The Guti were finally expelled from Mesopotamia by the Sumerians of Erech (c. 2100), but it was left to the kings of Ur's famous third dynasty to re-establish the Sargonoid frontiers and write the final chapter of the Sumerian History. The dynasty lasted through the twenty first century at the close of which the armies of Ur were overthrown by the Elamites and Amorites (?).

This passage does not feature individual people performing specific actions, but it is still united by character and action. Instead of Hermione at breakfast, we have the Sumarians in Mesopotamia. Instead of retrieving a message from an owl, the conquer the Guti, but then are conquered by the Elamites and Amorites. The important thing is that the statements in a narrative are not related as premises and conclusion. Instead, they are all events which are united common characters acting in specific times and places.

### *Socratic Elenchus*

Now we will consider one example of a collection of sentences that *is* an argument: the Socratic elenchus.

The Greek philosopher Socrates was known for his odd style of arguing with his interlocutors. While Socrates himself left no known writings, his students Plato and Xenophon wrote many dialogues (essentially, lengthy academic plays) in which Socrates was a main character. The character Socrates and the actual person Socrates are thought to have much in common (although this is still an object of much academic debate).

In many of the dialogues, Socrates engages his interlocutors (that is, his conversational and argumentative partners) in what is called the **SOCRATIC ELENCHUS** in order to demonstrate that their previously-held beliefs were false or contradictory. The elenchus is a style of argumentation in which two statements (usually definitions) are shown to be contradictory by means of critical questioning, so an alternative is sought.

For example, Socrates discusses the nature of piety with Euthyphro and asks Euthyphro whether pious actions are loved by the gods. Euthyphro, of course, answers in the affirmative. Socrates then asks whether pious actions are *made pious* by the gods' love or if the gods just so happen to love piety. Euthyphro is unsure, since he is not certain what *makes* pious actions pious in the first place. If the gods *make* those actions pious, Socrates reasons, then they could just as easily have made any actions whatsoever pious. However, if the gods simply love piety as it is, then this seems to detract from the power of the gods. After all, the gods would have no say in what they find good!

This is just one example of the Socratic elenchus.

### *Key Terms*

1. **Argument**
2. **Canonical form**
3. **Conclusion**
4. **Conclusion indicator**
5. **Content neutrality**
6. **Critical thinker**
7. **Critical thinking**
8. **Explanation**
9. **Expository passage**
10. **Formal logic**
11. **Inference**
12. **Informal logic**
13. **Logic**
14. **Metacognition**
15. **Metareasoning**
16. **Narrative**
17. **Practical argument**
18. **Premise**
19. **Premise indicator**
20. **Reason**
21. **Rhetoric**
22. **Simple statement of belief**
23. **Socratic elenchus**
24. **Statement**
25. **Target proposition**

## 2

# *The Basics of Evaluating Argument*

### *Two Ways an Argument Can Go Wrong*

Arguments are supposed to lead us to the truth but they don't always succeed. There are two ways that arguments can fail to lead us to true conclusions. First, they can simply start off with false premises. Consider the following argument:

- (A1) It is raining heavily.
- (A2) If it is raining heavily, then you should take an umbrella.
- (A3)  $\therefore$  So, you should take an umbrella.

If premise (1) is false—if it is sunny outside—then the argument gives you no reason to carry an umbrella. The argument has failed its job. Premise (2) could also be false: Even if it is raining outside, you might not need an umbrella. You might wear a rain poncho or keep to covered walkways and still avoid getting soaked. Again, the argument fails because a premise is false. An argument with false premises can not lead us to a true conclusion.

Even if an argument has all true premises, there is still a second way it can fail. Consider another example:

- (B1) You are reading this book.
- (B2) Most people who read this book are logic students.
- (B3)  $\therefore$  You are a logic student.

This is not a terrible argument. The premises are true. Most people who read this book *are* logic students. Yet, it is possible for someone besides a logic student to read this book. If your friend who is not currently in a logic class read this book, they would not immediately become a logic student. So the premises of this argument, even though they are true, do not guarantee the truth of the conclusion. The evidential support between premises and conclusion is not a guarantee. This criterion is about the *structure* of the argument, and how the premises and the conclusion are related to one another. There are better and worse ways that the premises of an argument

can supply evidential support to its conclusion. Compare the example above with this one:

- (C1) You are reading this book.
- (C2) At least one person reading this book is a professional surfer.
- (C3)  $\therefore$  You are a professional surfer.

This argument should strike you as substantially worse than the previous one, even if you really are a professional surfer! Suppose the premises are both true. It still would seem pretty unlikely that the premises are very good reason to think that the conclusion is true. Just because there's at least one professional surfer who read this book, it doesn't follow that that person is you. Even though both of these arguments fail to guarantee their conclusions, one does seem better than the other. We'll be discussing why some arguments that fail this second criterion may still be worthwhile arguments.

To sum up, for any argument there are two ways that it could fail. First, one or more of the premises might be false. Second, the premises might fail to support the conclusion. Even if the premises were true, the form of the argument might be weak, meaning that there is little to no evidential support from premises to conclusion.

### *Validity and Soundness*

In logic, we are mostly concerned with evaluating the quality of inferences, not the truth of the premises. The truth of various premises will be a matter of whatever specific topic we are arguing about and since logic is content neutral we will also remain neutral.

The strongest possible evidential support would be for the premises to somehow force the conclusion to be true. This kind of inference is called **VALID**. Let's make this notion a bit more precise:

An argument is valid if and only if it is impossible for the premises to be true and the conclusion false.

The important thing to see is that the definition is about what *would* happen if the premises were true. It doesn't state that the premises actually *are* true. This is why our definition is about what is possible or impossible. The argument is valid if, when you imagine the premises are true, you are somehow forced into imagining that the conclusion is also true. Consider the argument in Figure ??

- (D1) Lady Gaga is from Mars.
- (D2)  $\therefore$  Lady Gaga is from the fourth planet from our sun.

The American pop star Lady Gaga is not from Mars. (She's from New York City.) Nevertheless, if you grant that she is from Mars, you *also* have to



grant that she is from the fourth planet from our sun, because Mars simply is the fourth planet from our sun. Therefore this argument is valid.

This way of understanding validity is based on what you can imagine, but not everyone is convinced that the imagination is a reliable tool in logic. That is why definitions like ?? and ?? talk about what is necessary or impossible. If the premises are true, the conclusion necessarily must be true. Alternately, it is impossible for the premises to be true and the conclusion false. The idea here is that instead of talking about the imagination, we will just talk about what can or cannot happen at the same time. The fundamental notion of validity remains the same, however: the truth of the premises would simply guarantee the truth of conclusion.

So, assessing validity means wondering about whether the conclusion would be true *if* the premises were true. This means that valid arguments can have false conclusions. This is important to keep in mind because people naturally tend to think that any argument must be good if they agree with the conclusion. And the more passionately people believe in the conclusion, the more likely we are to think that any argument for it must be brilliant. Conversely, if the conclusion is something we don't believe in, we naturally tend to think the argument is poor. And the more we don't like the conclusion, the less likely we are to like the argument.

But this is not the correct way to evaluate inferences at all. The quality of the inference is entirely independent of the truth of the conclusion. You can have great arguments for false conclusions and horrible arguments for true conclusions. We have trouble seeing this because of biases built deep in the way we think called "cognitive biases." A **COGNITIVE BIAS** is a habit of reasoning that can be dysfunctional in certain circumstances. Generally these biases developed for a reason, so they serve us well in many or most circumstances. But cognitive biases also systematically distort our reasoning in other circumstances, so we must be on guard against them.

There is a particular cognitive bias that makes it hard for us to recognize when a poor argument is being given for a conclusion we agree with. It is called "confirmation bias" and it is in many ways the mother of all cognitive biases. **CONFIRMATION BIAS** is the tendency to discount or ignore evidence and arguments that contradict one's current beliefs. It really pervades all of our thinking, right down to our perceptions. We'll discuss confirmation bias in more detail in chapter 6.

Because of confirmation bias, we need to train ourselves to recognize valid arguments for conclusions we think are false. Remember, an argument is valid if it is impossible for the premises to be true and the conclusion false. This means that you can have valid arguments with false conclusions, they just have to also have false premises. Consider the example in Figure ??

(E1) Oranges are either fruits or musical instruments.

(E2) Oranges are not fruits.

(E3)  $\therefore$  Oranges are musical instruments.

The conclusion of this argument is ridiculous. Nevertheless, it follows validly from the premises. This is a valid argument. *If* both premises were true, *then* the conclusion would necessarily be true.

This shows that a valid argument does not need to have true premises or a true conclusion. Conversely, having true premises and a true conclusion is not enough to make an argument valid. Consider the example in Figure ??

(F1) London is in England.

(F2) Beijing is in China.

(F3)  $\therefore$  Paris is in France.

The premises and conclusion of this argument are, as a matter of fact, all true. This is a terrible argument, however, because the premises have nothing to do with the conclusion. Imagine what would happen if Paris declared independence from the rest of France. Then the conclusion would be false, even though the premises would both still be true. Thus, it is *logically possible* for the premises of this argument to be true and the conclusion false. The argument is not valid. If an argument is not valid, it is called **INVALID**. As we shall see, this term is a little misleading, because less than perfect arguments can be very useful. But before we do that, we need to look more at the concept of validity.

In general, then, the *actual* truth or falsity of the premises, if known, do not tell you whether or not an inference is valid. There is one exception: when the premises are true and the conclusion is false, the inference cannot be valid, because valid reasoning can only yield a true conclusion when beginning from true premises.

Figure ?? has another invalid argument:

(G1) All dogs are mammals.

(G2) All dogs are animals.

(G3)  $\therefore$  All animals are mammals.

In this case, we can see that the argument is invalid by looking at the truth of the premises and conclusion. We know the premises are true. We know that the conclusion is false. This is the one circumstance that a valid argument is supposed to make impossible.

Some invalid arguments are hard to detect because they resemble valid arguments. Consider the one in Figure ??

(H1) An economic stimulus package will allow the U.S. to avoid a depression.

(H2) There is no economic stimulus package

(H3)  $\therefore$  [3] The U.S. will go into a depression.

This reasoning is not valid since the premises do not *definitively* support the conclusion. To see this, assume that the premises are true and then ask, “Is it possible that the conclusion could be false in such a situation?”. There is no inconsistency in taking the premises to be true without taking the conclusion to be true. The first premise says that the stimulus package will allow the U.S. to avoid a depression, but it does not say that a stimulus package is the *only* way to avoid a depression. Thus, the mere fact that there is no stimulus package does not necessarily mean that a depression will occur.

When an argument resembles a good argument but is actually a bad one, we say it is a **FALLACY**. Fallacies are similar to cognitive biases, in that they are ways our reasoning can go wrong. Fallacies, however, are always mistakes you can explicitly lay out as arguments in canonical form, as above.

Here is another, trickier, example. I will give it first in ordinary language.

*A pundit is speaking on a cable news show* If the U.S. economy were in recession and inflation were running at more than 4%, then the value of the U.S. dollar would be falling against other major currencies. But this is not happening — the dollar continues to be strong. So, the U.S. is not in recession.

The conclusion is “The U.S. economy is not in recession.” If we put the argument in canonical form, it looks like figure ??

- (I1) If the U.S. were in a recession with more than 4% inflation, then the dollar would be falling.
- (I2) The dollar is not falling.
- (I3)  $\therefore$  The U.S. is not in a recession.

The conclusion does not follow necessarily from the premises. It does follow necessarily from the premises that (i) the U.S. economy is not in recession or (ii) inflation is running at more than 4%, but they do not guarantee (i) in particular, which is the conclusion. For all the premises say, it is possible that the U.S. economy is in recession but inflation is less than 4%. So, the inference does not *necessarily* establish that the U.S. is not in recession. A parallel inference would be “Jack needs eggs and milk to make an omelet. He can’t make an omelet. So, he doesn’t have eggs.”.

If an argument is not only valid, but also has true premises, we call it **SOUND**. “Sound” is the highest compliment you can pay an argument. If logic is the study of virtue in argument, sound arguments are the most virtuous. We said in Section 2 that there were two ways an argument could go wrong, either by having false premises or weak inferences. Sound arguments have true premises and undeniable inferences. If someone gives a sound argument in a conversation, you have to believe the conclusion, or else you are irrational.

- |  |  |
|--|--|
| (J1) Socrates is a person.                         | (K1) Socrates is a person.                       |
| (J2) All people are carrots.                       | (K2) All people are mortal.                      |
| (J3) $\therefore$ Therefore, Socrates is a carrot. | (K3) $\therefore$ Therefore, Socrates is mortal. |

**Valid, but not sound****Valid and sound**

The argument on the left in Figure 3.1 is valid, but not sound. The argument on the right is both valid and sound.

Both arguments have the exact same form. They say that a thing belongs to a general category and everything in that category has a certain property, so the thing has that property. Because the form is the same, it is the same valid inference each time. The difference in the arguments is not the validity of the inference, but the truth of the second premise. People are not carrots, therefore the argument on the left is not sound. People are mortal, so the argument on the right is sound.

Often it is easy to tell the difference between validity and soundness if you are using completely silly examples. Things become more complicated with false premises that you might be tempted to believe, as in the argument in Figure ??.

- (L1) Every Irishman drinks Guinness.
- (L2) Smith is an Irishman.
- (L3)  $\therefore$  Smith drinks Guinness.

You might have a general sense that the argument in Figure ?? is bad—you shouldn't assume that someone drinks Guinness just because they are Irish. But the argument is completely valid (at least when it is expressed this way.) The inference here is the same as it was in the previous two arguments. The problem is the first premise. Not all Irishmen drink Guinness, but if they did, and Smith was an Irishman, he would drink Guinness.

The important thing to remember is that validity is not about the actual truth or falsity of the statements in the argument. Instead, it is about the way the premises and conclusion are put together. It is really about the *form* of the argument. A valid argument has perfect logical form. The premises and conclusion have been put together so that the truth of the premises is incompatible with the falsity of the conclusion.

A general trick for determining whether an argument is valid is to try to come up with just one way in which the premises could be true but the conclusion false. If you can think of one (just one! anything at all! but no violating the laws of physics!), the reasoning is *invalid*.

### *Strong, Cogent, Deductive, Inductive*

We have just seen that sound arguments are the very best arguments. Unfortunately, sound arguments are really hard to come by, and when you do find them, they often only prove things that were already quite obvious, like that Socrates (a dead man) is mortal. Fortunately, arguments can still be worthwhile, even if they are not sound. Consider this one:

- (M1) In January 1997, it rained in San Diego.
- (M2) In January 1998, it rained in San Diego.
- (M3) In January 1999, it rained in San Diego.
- (M4)  $\therefore$  It rains every January in San Diego.

This argument is not valid, because the conclusion could be false even though the premises are true. It is possible, although unlikely, that it will fail to rain next January in San Diego. Moreover, we know that the weather can be fickle. No amount of evidence should convince us that it rains there *every* January. Who is to say that some year will not be a freakish year in which there is no rain in January in San Diego? Even a single counterexample is enough to make the conclusion of the argument false.

Still, this argument is pretty good. Certainly, the argument could be made stronger by adding additional premises: In January 2000, it rained in San Diego. In January 2001. . . and so on. Regardless of how many premises we add, however, the argument will still not be deductively valid. Instead of being valid, this argument is strong. An argument is **STRONG** if the premises would make the conclusion more likely, were they true. In a strong argument, the premises don't guarantee the truth of the conclusion, but they do make it a good bet. If an argument is strong, and it has true premises, we say that it is **COGENT**. Cogency is the equivalent of soundness in strong arguments. If an inference is neither valid, nor strong, we say it is **WEAK**. In a weak argument, the premises would not even make the conclusion likely, even if they were true.

You may have noticed that the word “likely” is a little vague. How likely do the premises have to make the conclusion before we can count the argument as strong? The answer is a very unsatisfying “it depends.” It depends on what is at stake in the decision to believe the conclusion. What happens if you are wrong? What happens if you are right? The phrase “make the conclusion a good bet” is really quite apt. Whether something is a good bet depends a lot on how much money is at stake and how much you are willing to lose. Sometimes people feel comfortable taking a bet that has a 50% chance of doubling their money, sometimes they don't.

The vagueness of the word “likely” brings out an interesting feature of strong arguments: some strong arguments are stronger than others. The argument about rain in San Diego, above, has three premises referring to three previous Januaries. The argument is pretty strong, but it can

become stronger if we go back farther into the past, and find more years where it rains in January. The more evidence we have, the better a bet the conclusion is. Validity is not like this. Validity is a black-or-white matter. You either have it, and you're perfect, or you don't, and you're nothing. There is no point in adding premises to an argument that is already valid.

Arguments that are valid, or at least try to be, are called **DEDUCTIVE**, and people who attempt to argue using valid arguments are said to be arguing *deductively*. The notion of validity we are using here is, in fact, sometimes called *deductive validity*. Deductive argumentation is difficult, because, as we said, in the real world sound arguments are hard to come by, and people don't always recognize them as sound when they find them. Arguments that purport to merely be strong rather than valid are called **INDUCTIVE**. The most common kind of inductive argument includes arguments like the one above about rain in San Diego, which generalize from many cases to a conclusion about all cases.

Deduction is possible in only a few contexts. You need to have clear, fixed meanings for all of your terms and rules that are universal and have no exceptions. One can find situations like this if you are dealing with things like legal codes, mathematical systems or logical puzzles. One can also create, as it were, a context where deduction is possible by imagining a universal, exceptionless rule, even if you know that no such rule exists in reality. In the example above about rain in San Diego, we can change the argument from inductive to deductive by adding a universal, exceptionless premise like "It always rains in January in San Diego." This premise is unlikely to be true, but it can make the inference valid. (For more about trade offs between the validity of the inference and the truth of the premise, see the chapter on incomplete arguments in the complete version of this text.

Here is an example in which the context is an artificial code — the tax code:

*From a the legal code posted on a government website* A tax credit for energy-efficient home improvement is available at 30% of the cost, up to \$1,500 total, in 2009 & 2010, ONLY for existing homes, NOT new construction, that are your "principal residence" for Windows and Doors (including sliding glass doors, garage doors, storm doors and storm windows), Insulation, Roofs (Metal and Asphalt), HVAC: Central Air Conditioners, Air Source Heat Pumps, Furnaces and Boilers, Water Heaters: Gas, Oil, & Propane Water Heaters, Electric Heat Pump Water Heaters, Biomass Stoves.

This rule describes the conditions under which a person can or cannot take a certain tax credit. Such a rule can be used to reach a valid conclusion that the tax credit can or cannot be taken.

As another example of an inference in an artificial situation with limited and clearly defined options, consider a Sudoku puzzle. The rules of Sudoku are that each cell contains a single number from 1 to 9, and each row, each

column and each 9-cell square contain one occurrence of each number from 1 to 9. Consider the following partially completed board:

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

The following inference shows that, in the first column, a 9 must be entered below the 7:

The 9 in the first column must go in one of the open cells in the column. It cannot go in the third cell in the column, because there is already a 9 in that 9-cell square. It cannot go in the eighth or ninth cell because each of these rows already contains a 9, and a row cannot contain two occurrences of the same number. Therefore, since there must be a 9 somewhere in this column, it must be entered in the seventh cell, below the 7.

The reasoning in this inference is valid: if the premises are true, then the conclusion must be true. Logic puzzles of all sorts operate by artificially restricting the available options in various ways. This then means that each conclusion arrived at (assuming the reasoning is correct) is necessarily true.

One can also create a context where deduction is possible by imagining a rule that holds without exception. This can be done with respect to any subject matter at all. Speakers often exaggerate the connecting premise in order to ensure that the justificatory or explanatory power of the inference is as strong as possible. Consider Smith's words in the following passage:

**Smith:** I'm going to have some excellent pizza this evening.

**Jones:** I'm glad to hear it. How do you know?

**Smith:** I'm going to Adriatico's. They always make a great pizza.

Here, Smith justifies his belief that the pizza will be excellent — it comes from Adriatico's, where the pizza, he claims, is *always* great: in the past, present and future.

As stated by Smith, the inference that the pizza will be great this evening is valid. However, making the inference valid in this way often means

making the general premise false: it's not likely that the pizza is great *every single* time; Smith is overstating the case for emphasis. Note that Smith does not need to use a universal proposition in order to convince Jones that the pizza will *very likely* be good. The inference to the conclusion would be strong (though not valid) if he had said that the pizza is "almost always" great, or that the pizza has been great on all of the many occasions he has been at that restaurant in the past. The strength of the inference would fall to some extent—it would not be guaranteed to be great this evening—but a slightly weaker inference seems appropriate, given that sometimes things go contrary to expectation.

Sometimes the laws of nature make constructing contexts for valid arguments more reasonable. Now consider the following passage, which involves a scientific law:

Jack is about to let go of Jim's leash. The operation of gravity makes all unsupported objects near the Earth's surface fall toward the center of the Earth. Nothing stands in the way. Therefore, Jim's leash will fall.

(Or, as Spock said in a Star Trek episode, "If I let go of a hammer on a planet that has a positive gravity, I need not see it fall to know that it has in fact fallen.") The inference above is represented in canonical form as follows:

- (N1) Jack is about to let go of Jim's leash.
- (N2) The operation of gravity makes all unsupported objects near the Earth's surface fall toward the center of the Earth.
- (N3) Nothing stands in the way of the leash falling.
- (N4)  $\therefore$  Jim's leash will fall toward the center of the Earth.

As stated, this argument is valid. That is, if you pretend that they are true or accept them "for the sake of argument", you would *necessarily* also accept the conclusion. Or, to put it another way, there is no way in which you could hold the premises to be true and the conclusion false.

Although this argument is valid, it involves idealizing assumptions similar to the ones we saw in the pizza example. P<sub>2</sub> states a physical law which is about as well confirmed as any statement about the world around us you care to name. However, physical laws make assumptions about the situations they apply to—they typically neglect things like wind resistance. In this case, the idealizing assumption is just that nothing stands in the way of the leash falling. This can be checked just by looking, but this check can go wrong. Perhaps there is an invisible pillar underneath Jack's hand? Perhaps a huge gust of wind will come? These events are much less likely than Adriatico's making a lousy pizza, but they are still possible.

Thus we see that using scientific laws to create a context where deductive validity is possible is a much safer bet than simply asserting whatever exceptionless rule pops into your head. However, it still involves improving



the quality of the inference by introducing premises that are less likely to be true.

So deduction is possible in artificial contexts like logical puzzles and legal codes. It is also possible in cases where we make idealizing assumptions or imagine exceptionless rules. The rest of the time we are dealing with induction. When we do induction, we try for strong inferences, where the premises, assuming they are true, would make the truth of the conclusion very likely, though not necessary. Consider the two arguments in Figure 2.1

Figure 2.1: Neither argument is valid, but one is strong and one is weak

(O1) 92% of Republicans from Texas voted for Bush in 2000.	(P1) Just over half of drivers are female.
(O2) Jack is a Republican from Texas.	(P2) There's a person driving the car that just cut me off.
(O3) ∴ Jack voted for Bush.	(P3) ∴ The person driving the car that just cut me off is female.
A <b>strong</b> argument	A <b>weak</b> argument

Note that the premises in neither inference *guarantee* the truth of the conclusion. For all the premises in the first one say, Jack could be one of the 8% of Republicans from Texas who did not vote for Bush; perhaps, for example, Jack soured on Bush, but not on Republicans in general, when Bush served as governor. Likewise for the second; the driver could be one of the 49%.

So, neither inference is valid. But there is a big difference between how much support the premises, if true, would give to the conclusion in the first and how much they would in the second. The premises in the first, assuming they are true, would provide very strong reasons to accept the conclusion. This, however, is not the case with the second: if the premises in it were true then they would give only weak reasons for believing the conclusion. thus, the first is strong while the second is weak.

As we said earlier, there are only two options with respect to validity—valid or not valid. On the other hand, strength comes in degrees, and sometimes arguments will have percentages that will enable you to exactly quantify their strength, as in the two examples in Figure 2.1.

However, even where the degree of support is made explicit by a percentage there is no firm way to say at what degree of support an inference can be classified as strong and below which it is weak. In other words, it is difficult to say whether or not a conclusion is *very likely* to be true. For example, In the inference about whether Jack, a Texas Republican, voted for Bush. If 92% of Texas Republicans voted for Bush, the conclusion, if the premises are

granted, would very probably be true. But what if the number were 85%? Or 75%? Or 65%? Would the conclusion very likely be true? Similarly, the second inference involves a percentage greater than 50%, but this does not seem sufficient. At what point, however, would it be sufficient?

In order to answer this question, go back to basics and ask yourself: "If I accept the truth of the premises, would I then have sufficient reason to believe the conclusion?". If you would not feel safe in adopting the conclusion as a belief as a result of the inference, then you think it is weak, that is, you do not think the premises give sufficient support to the conclusion.

Note that the same inference might be weak in one context but strong in another, because the degree of support needed changes. For example, if you merely have a deposit to make, you might accept that the bank is open on Saturday based on your memory of having gone to the bank on Saturday at some time in the past. If, on the other hand, you have a vital mortgage payment to make, you might not consider your memory sufficient justification. Instead, you will want to call up the bank and increase your level of confidence in the belief that it will be open on Saturday.<sup>1</sup>

Most inferences (if successful) are strong rather than valid. This is because they deal with situations which are in some way open-ended or where our knowledge is not precise. In the example of Jack voting for Bush, we know only that 92% of Republicans voted for Bush, and so there is no definitive connection between being a Texas Republican and voting for Bush. Further, we have only statistical information to go on. This statistical information was based on polling or surveying a sample of Texas voters and so is itself subject to error (as is discussed in the chapter on induction in the complete version of this text. A more precise version of the premise might be "92%  $\pm$  3% of Texas Republicans voted for Bush".

At the risk of redundancy, let's consider a variety of examples of valid, strong and weak inferences, presented in standard form.

(Q1) David Duchovny weighs more than 200 pounds.

(Q2)  $\therefore$  David Duchovny weighs more than 150 pounds.

The inference here is valid. It is valid because of the number system (here applied to weight): 200 is more than 150. It might be false, as a matter of fact, that David Duchovny weighs more than 200 pounds, and false, as a matter of fact, that David Duchovny weighs more than 150 pounds. But if you *suppose* or *grant* or *imagine* that David Duchovny weighs more than 200 pounds, it would then *have* to be true that David Duchovny weighs more than 150 pounds. Next:

(R1) Armistice Day is November 11th, each year.

(R2) Halloween is October 31st, each year.

(R3)  $\therefore$  Armistice Day is later than Halloween, each year.

<sup>1</sup> This kind of context-sensitivity is called *contextualism* in epistemology. For more information, see e.g. Jason Stanley's book *Knowledge and Practical Interests*.

This inference is valid. It is valid because of order of the months in the Gregorian calendar and the placement of the New Year in this system. Next:

- (S1) All people are mortal.
- (S2) Professor Pappas is a person.
- (S3)  $\therefore$  Professor Pappas is mortal.

As written, this inference is valid. If you accept for the sake of argument that all men are mortal (as the first premise says) and likewise that Professor Pappas is a man (as the second premise says), then you would have to accept also that Professor Pappas is mortal (as the conclusion says). You could not consistently both (i) affirm that all men are mortal and that Professor Pappas is a man and (ii) deny that Professor Pappas is mortal. If a person accepted these premises but denied the conclusion, that person would be making a mistake in logic.

This inference's validity is due to the fact that the first premise uses the word "all". You might, however, wonder whether or not this premise is true, given that we believe it to be true only on our experience of men *in the past*. This might be a case of over-stating a premise, which we mentioned earlier. Next:

- (T1) In 1933, it rained in Columbus, Ohio on 175 days.
- (T2) In 1934, it rained in Columbus, Ohio on 177 days.
- (T3) In 1935, it rained in Columbus, Ohio on 171 days.
- (T4)  $\therefore$  In 1936, it rained in Columbus, Ohio on at least 150 days.

This inference is strong. The premises establish a record of days of rainfall that is well above 150. It is possible, however, that 1936 was exceptionally dry, and this possibility means that the inference does not achieve validity. Next:

- (U1) The Bible says that (male) homosexuality is an abomination.
- (U2)  $\therefore$  (Male) homosexuality is an abomination.

This inference is an appeal to a source. In brief, you should think about whether the source is reliable, biased, or whether the claim is consistent with what other authorities on the subject say. You should apply all these criteria to this argument for yourself. You should ask what issues, if any, the Bible is reliable on. If you believe humans had any role in writing the Bible, you can ask about what biases and agendas they might have had. And you can think about what other sources—religious texts or moral experts—say on this issue. You can certainly find many who disagree. We can also evaluate the context of the specific verse (Leviticus 18:22) that the premise refers to. It is important to note that here we are not evaluating the inference by taking a position on whether the premise is true (it is) or conclusion is true (it isn't). When evaluating an inference we are evaluating

the *degree of evidential support* that the premises confer on the conclusion. Perhaps an appeal to a religious text is very good (if defeasible) evidence for some conclusion, perhaps it isn't.

(V1) Some professional philosophers published books in 2007.

(V2) Some books published in 2007 sold more than 100,000 copies.

(V3)  $\therefore$  Some professional philosophers published books in 2007 that sold more than 100,000 copies.

This reasoning is weak. Both premises use the word “some” which doesn't tell you much about how many professional philosophers published books and how many books sold more than 100,000 copies in 2007. This means that you cannot be confident that even one professional philosopher sold more than 100,000 copies. Next:

(W1) Lots of Russians prefer vodka to bourbon.

(W2)  $\therefore$  George Bush was the President of the United States in 2006.

No one would sincerely make an inference like this. It is presented here as an example only: it is clearly weak. It's hard to see how the premise justifies the conclusion to any extent at all.

To sum up this section, we have seen that there are two styles of reasoning, deductive and inductive. The former tries to use valid arguments, while the latter contents itself to give arguments that are strong.

### *Key Terms*

- |                             |                   |
|-----------------------------|-------------------|
| 1. <b>Cogent</b>            | 7. <b>Invalid</b> |
| 2. <b>Cognitive Bias</b>    | 8. <b>Sound</b>   |
| 3. <b>Confirmation bias</b> | 9. <b>Strong</b>  |
| 4. <b>Deductive</b>         | 10. <b>Valid</b>  |
| 5. <b>Fallacy</b>           | 11. <b>Weak</b>   |
| 6. <b>Inductive</b>         |                   |

### 3

## *What is Formal Logic?*

### *Formal as in Concerned with the Form of Things*

??

Many of the other chapters in this book will deal with formal logic. Formal logic is distinguished from other branches of logic by the way it achieves content neutrality. Back on page 24, we said that a distinctive feature of logic is that it is neutral about the content of the argument it evaluates. If a kind of argument is strong—say, a kind of statistical argument—it will be strong whether it is applied to sports, politics, science or whatever. Formal logic takes radical measures to ensure content neutrality: it removes the parts of a statement that tie it to particular objects in the world and replaces them with abstract symbols. (See page 25)

Consider the two arguments from Figure 3.1 again:

- |                                       |   |
|---------------------------------------|---|
| (A1) Socrates is a person.            | (B1) Socrates is a person.              |
| (A2) All persons are mortal.          | (B2) All people are carrots.            |
| (A3) $\therefore$ Socrates is mortal. | (B3) $\therefore$ Socrates is a carrot. |

Figure 3.1: Two examples of valid arguments.

These arguments are both valid. In each case, if the premises were true, the conclusion would have to be true. (In the case of the first argument, the premises are actually true, so the argument is sound, but that is not what we are concerned with right now.) What makes these arguments valid is that they are put together the right way. Another way of thinking about this is to say that they have the same logical form. Both arguments can be written like this:

- (C1)  $S$  is  $M$ .  
(C2) All  $M$  are  $P$ .  
(C3)  $\therefore S$  is  $P$ .

In both arguments  $S$  stands for Socrates and  $M$  stands for person. In the first argument,  $P$  stands for mortal; in the second,  $P$  stands for carrot.

The letters ‘S’, ‘M’, and ‘P’ are variables. They are just like the variables you may have learned about in algebra class. In algebra, you had equations like  $y = 2x + 3$ , where  $x$  and  $y$  were variables that could stand for any number. Just as  $x$  could stand for any number in algebra, ‘S’ can stand for any name in logic. In fact, this is one of the original uses of variables. Long before variables were used to stand for numbers in algebra, they were used to stand for classes of things, like people or carrots, by Aristotle in his book the (*?*). Around the same time in India, the grammarian and linguist Pāṇini was also using variables to represent possible sounds that could be used in different forms of a word (*?*). Both thinkers introduce their variables fairly causally, as if their readers would be familiar with the idea, so it may be that people prior to them actually invented the variable!

Whoever invented it, the variable was one of the most important conceptual innovations in human history, right up there with the invention of zero or alphabetic writing. The importance of the variable for the history of mathematics is obvious. But it was also incredibly important in one of its original fields of application, logic. For one thing, it allows logicians to be more content neutral. We can set aside any associations we have with people, or carrots, or whatever, when we are analyzing an argument. More importantly, once we set aside content in this way, we discover that something incredibly powerful is left over, the logical structure of the sentence itself. This is what we investigate when we study formal logic. In the case of the two arguments above, identifying the logical structure of statements reveals not only that the two arguments have the same logical form, but they have an impeccable logical form. Both arguments are valid, and any other arguments that have this form will be valid.

When Aristotle introduced the variable to the study of logic he used it the way we did in the argument above. His variables stood for names and categories in simple two-premise arguments called syllogisms. The system of logic Aristotle outlined became the dominant logic in the Western world for more than two millennia. It was studied and elaborated on by philosophers and logicians from Baghdad to Paris. The thinkers that carried on Aristotelian tradition were divided by language and religion. They were pagans, Muslims, Jews, and Christians writing typically in Greek, Latin or Arabic. But they were all united by the sense that the project Aristotle had initiated allowed them to see something profound about the nature of reality. They were looking at abstract structures which somehow seemed to be at the foundation of things. As the philosopher and historian of logic Catarina Dutilh Novaes points out, the logic that the thinkers of all these religious traditions were pursuing was formal in that it concerned the *forms* of things (*?*). As formal logic evolved, however, the idea of being “formal” would take on an additional meaning.

## *Formal as in Strictly Following Rules*

??

Despite its historical importance, Aristotelean logic has largely been superseded. Starting in the 19th century people learned to do more than simply replace categories with variables. They learned to replicate the whole structure of sentences with a formal system that brought out all sorts of features of the logical form of arguments. The result was the creation of entire artificial languages. An **ARTIFICIAL LANGUAGE** is a language that was consciously developed by identifiable individuals for some purpose. Esperanto, for instance, is an artificial language developed by Ludwig Lazarus Zamenhof in the 19th century with the hope of promoting world peace by creating a common language for all. J.R.R. Tolkien invented several languages to flesh out the fictional world of his fantasy novels, and even created timelines for their evolution. For Tolkien, the creation of languages was an art form in itself, “An art for which life is not long enough, indeed: the construction of imaginary languages in full or outline for amusement, for the pleasure of the constructor or even conceivably of any critic that might occur”<sup>1</sup>. And it is an art that is really beginning to catch on, especially with Hollywood commissioning languages to be constructed for blockbuster films.

Artificial languages contrast with **NATURAL LANGUAGES**, which develop spontaneously and are learned by infants as their first language. Natural languages include all the well-known languages spoken around the world, like English or Japanese or Arabic. It also includes more recently developed languages and evolved spontaneously amongst groups of people. For instance, deaf children will often spontaneously develop their own sign language. This phenomenon was important for the development of American Sign Language (ASL) and is part of why ASL counts as a *natural* language. For similar reasons Nicaraguan Sign Language counts as a natural language, even though it emerged very recently—in the late 1970s and 80s—after the Sandinista government set up schools for the deaf in Nicaragua for the first time. Natural languages can also develop by creolization, when languages merge and children grow up speaking the merged language as their first language. Haitian Creole is one of the most famous examples of this.

The languages developed by logicians are artificial, not natural. Their goal is not to promote global harmony, like Zamenhof’s Esperanto. Nor are they creating art for art’s sake, as Tolkien was, although logical languages can have a great deal of beauty. When the languages first started being developed in the late 19th and early 20th centuries, the goal was, in fact, to have a logically pure language, free of the irrationalities the plague natural languages. More specifically, they had two distinct goals: first, remove all ambiguity and vagueness, and second, to make the logical structure of the language immediately apparent, so that the language wore its logical

structure on its face, as it were. If such a language could be developed, it would help us solve all kinds of problems. The logician and philosopher Rudolf Carnap, for instance, felt that the right artificial language could simply make philosophical problems disappear (?).

The languages developed by logicians in the late 19th and early 20th centuries got labeled formal languages, in part because the logicians in question were working in the tradition of formal logic that was already established. A shift began to happen here with the meaning of formal, however, a change which is well documented by Dutilh Novaes <sup>(2)</sup>. Logicians began to hope that the languages that were being developed were so logical that everything about them could be characterized by a machine. A machine could be used to create sentences in this language, and then again to identify all the valid arguments in this language. This brings out another sense of the word “formal.” As Dutilh Novaes puts it <sup>(3)</sup> instead of being “formal” in the sense of concerning the forms of things, logic was formal in the sense that it followed rules perfectly precisely. You might compare this to the way a “formal hearing” in a court of law follows the rule of law to the letter.

For our purposes we will say that the core idea of a **FORMAL LANGUAGE** is that it is an *artificial language designed to bring out the logical structure of ideas* and remove all the ambiguity and vagueness that plague natural languages like English. We will further add that sometimes, formal languages are languages that can be implemented by a machine. Creating formal languages always involves all kinds of trade offs. On the one hand, we are trying to create a language that makes a logical structure clear and obvious. This will require simplifying things, removing excess baggage from the language. On the other hand, we want to make the language perfectly precise, free of vagueness and ambiguity. This will mean adding complexity to the language. The other thing was that it was very important for the people developing these languages that you be able to prove all the truths of mathematics in them. This meant that the languages had to have a certain scope and certain limitations.

This was a trade off no logician was ever able to get perfectly correct, because, as it turns out, a logically pure language is impossible. No formal language can express everything that a natural language can. Logicians became convinced of this, naturally enough, because of a pair of logical proofs. In 1931, the logician Kurt Gödel showed that you couldn’t do all of mathematics in a consistent logical system, which was enough to persuade most of the logicians engaged in the project to drop it. There is a more general problem with the idea of a purely logical language, though, which is that many of the features logicians were trying to remove from language were actually necessary to make it function. Arika Okrent puts the point quite well. For Okrent, the failure of artificial languages is precisely what illuminates the virtues of natural language.



[By studying artificial languages we] gain a deeper appreciation of natural language and the messy qualities that give it so much flexibility and power and that a simple communication device. The ambiguity and lack of precision allow to serve as a instrument of thought *formation*, of experimentation and discovery. We don't know exactly what we mean before we speak; we can figure it out as we go along. We can talk just to talk, to be social, to feel connected, to participate. At the same time natural language still works as an instrument of thought transmission, one that can be *made* extremely precise and reliable when we need it to be, or left loose and sloppy when we can't spare the time or effort (?)

The languages developed in the late 19th and early 20th centuries had goals that were theoretical, rather than practical. They languages were meant to improve our understanding of the world for the sake of improving our understanding of the world. They failed at this theoretical goal, but they wound up having a practical spin-off of world-historical proportions, which is why formal logic is a thriving discipline to this day. Remember that in this period people started thinking of formal languages as languages that could be implemented mechanically. At first, the idea of a a mechanistic language was a metaphor. The rules that were being followed to the letter were to be followed by a human being actually writing down symbols. This human being was generally referred to as a “computer,” because they were computing things. The world changed when a logician named Alan Turing started using literal machines to be computers.

In the 1930s, Turing developed the idea of a reasoning machine that could compute any function. At first, this was just an abstract idea: it involved an infinite stretch of tape. But during World War II, Turing went to work the British code breaking effort at Bletchley Park. The Nazis encoded messages using a device called the Enigma Machine. The Allies had captured one, but since they settings on the machine were reshuffled for each message, it didn't do them much good. Turing, together with people like the mathematicians Gordon Welchman and Joan Clarke, managed to build another machine that could test Enigma settings rapidly to identify the configuration being used. People had made computing machines before, but now the science of logic was so much more advanced that they real power of mechanical computing could be exploited. The human computers became the fully programmable machines we know today, and the formal languages logicians created for theoretical reasons came the computer languages the world of the 21st century depends on. (All of this information, plus lots of fascinating pictures and diagrams, is available at [www.turing.org.uk](http://www.turing.org.uk).)

Part ?? of this book begins by exploring the world of Aristotelian logic, where logic is “formal” in the sense of being about the forms of things. Chapter 7 looks at the logical structure of the individual statements studied by the Aristotelian tradition. Chapter ?? then builds these into valid arguments. After we study Aristotelian logic, we will develop two formal languages, called SL and QL. Chapters 10 through ?? develop SL. In SL, the

smallest units are individual statements. Simple statements are represented as letters and connected with logical connectives like *and* and *not* to make more complex statements. Chapters ?? through ?? develop QL. In QL, the basic units are objects, properties of objects, and relations between objects. Part III of this book explores the world of Aristotelian logic. Chapter 7 looks at the logical structure of the individual statements studied by the Aristotelian tradition. Chapter ?? then builds these into valid arguments. Part ?? develops a full-blown formal language, called Sentential logic, or SL. In SL Simple statements are represented as letters and connected with logical connectives like *and* and *not* to make more complex statements. In this book we will be developing two formal languages, called SL and QL. Part ?? develops SL. In SL, the smallest units are individual statements. Simple statements are represented as letters and connected with logical connectives like *and* and *not* to make more complex statements. Part ?? develops QL. In QL, the basic units are objects, properties of objects, and relations between objects.

### *More Logical Notions for Formal Logic*

Part I covered the basic concepts you need to study any kind of logic. When we study formal logic, we will be interested in some additional logical concepts, which we will explain here.

#### *Truth values*

A truth value is the status of a statement as true or false. Thus the truth value of the sentence “All dogs are mammals” is “True,” while the truth value of “All dogs are reptiles” is false. More precisely, a **TRUTH VALUE** is the status of a statement with relationship to truth. We have to say this, because there are systems of logic that allow for truth values besides “true” and “false,” like “maybe true,” or “approximately true,” or “kinda sorta true.” For instance, some philosophers have claimed that the future is not yet determined. If they are right, then statements about *what will be the case* are not yet true or false. Some systems of logic accommodate this by having an additional truth value. Other formal languages, so-called paraconsistent logics, allow for statements that are both true *and* false. We won’t be dealing with those in this textbook, however. For our purposes, there are two truth values, “true” and “false,” and every statement has exactly one of these. Logical systems like ours are called **BIVALENT**.

#### *Tautology, contingent statement, contradiction*

In considering arguments formally, we care about what would be true *if* the premises were true. Generally, we are not concerned with the actual truth value of any particular statements— whether they are *actually* true or false.

Yet there are some statements that must be true, just as a matter of logic.

Consider these statements:

1. It is raining.
2. Either it is raining, or it is not.
3. It is both raining and not raining.

In order to know if statement 1 is true, you would need to look outside or check the weather channel. Logically speaking, it might be either true or false. Statements like this are called *contingent* statements.

Statement 2 is different. You do not need to look outside to know that it is true. Regardless of what the weather is like, it is either raining or not. If it is drizzling, you might describe it as partly raining or in a way raining and a way not raining. However, our assumption of bivalence means that we have to draw a line, and say at some point that it is raining. And if we have not crossed this line, it is not raining. Thus the statement “either it is raining or it is not” is always going to be true, no matter what is going on outside. A statement that has to be true, as a matter of logic is called a **TAUTOLOGY** or logical truth.

You do not need to check the weather to know about statement 3, either. It must be false, simply as a matter of logic. It might be raining here and not raining across town, it might be raining now but stop raining even as you read this, but it is impossible for it to be both raining and not raining here at this moment. The third statement is *logically false*; it is false regardless of what the world is like. A logically false statement is called a **CONTRADICTION**.

We have already said that a contingent statement is one that could be true, or could be false, as far as logic is concerned. To be more precise, we should define a **CONTINGENT STATEMENT** as a statement that is neither a tautology nor a contradiction. This allows us to avoid worrying about what it means for something to be logically possible. We can just piggyback on the idea of being logically necessary or logically impossible.

A statement might *always* be true and still be contingent. For instance, it may be the case that in no time in the history of the universe was there ever an elephant with tiger stripes. Elephants only ever evolved on Earth, and there was never any reason for them to evolve tiger stripes. The statement “Some elephants have tiger stripes,” is therefore always false. It is, however, still a contingent statement. The fact that it is always false is not a matter of logic. There is no contradiction in considering a possible world in which elephants evolved tiger stripes, perhaps to hide in really tall grass. The important question is whether the statement *must* be true, just on account of logic.

When you combine the idea of tautologies and contradictions with the notion of deductive validity, as we have defined it, you get some curious

results. For one thing, any argument with a tautology in the conclusion will be valid, even if the premises are not relevant to the conclusion. This argument, for instance, is valid.

(D1) There is coffee in the coffee pot.

(D2) There is a dragon playing bassoon on the armoire.

(D3)  $\therefore$  All bachelors are unmarried men.

The statement “All bachelors are unmarried men” is a tautology. No matter what happens in the world, all bachelors have to be unmarried men, because that is how the word “bachelor” is defined. But if the conclusion of the argument is a tautology, then there is no way that the premises could be true and the conclusion false. So the argument must be valid.

Even though it is valid, something seems really wrong with the argument above. The premises are not relevant to the conclusion. Each sentence is about something completely different. This notion of relevance, however, is something that we don’t have the ability to capture in the kind of simple logical systems we will be studying. The logical notion of validity we are using here will not capture everything we like about arguments.

Another curious result of our definition of validity is that any argument with a contradiction in the premises will also be valid. In our kind of logic, once you assert a contradiction, you can say anything you want. This is weird, because you wouldn’t ordinarily say someone who starts out with contradictory premises is arguing well. Nevertheless, an argument with contradictory premises is valid.

### *Logically Equivalent and Contradictory Pairs of Sentences*

We can also ask about the logical relations *between* two statements. For example:

(a) John went to the store after he washed the dishes.

(b) John washed the dishes before he went to the store.

These two statements are both contingent, since John might not have gone to the store or washed dishes at all. Yet they must have the same truth value. If either of the statements is true, then they both are; if either of the statements is false, then they both are. When two statements necessarily have the same truth value, we say that they are **LOGICALLY EQUIVALENT**.

On the other hand, if two sentences must have opposite truth values, we say that they are **CONTRADICTORIES**. Consider these two sentences

(a) Susan is taller than Monica.

(b) Susan is shorter or the same height as Monica.

One of these sentences must be true, and if one of the sentences is true, the other one is false. It is important to remember the difference between a single sentence that is a *contradiction* and a pair of sentences that are *contradictory*. A single sentence that is a contradiction is in conflict with itself, so it is never true. When a pair of sentences is contradictory, one must always be true and the other false.

### Consistency

Consider these two statements:

- (a) My only brother is taller than I am.
- (b) My only brother is shorter than I am.

Logic alone cannot tell us which, if either, of these statements is true. Yet we can say that *if* the first statement (a) is true, *then* the second statement (b) must be false. And if (b) is true, then (a) must be false. It cannot be the case that both of these statements are true. It is possible, however that both statements can be false. My only brother could be the same height as I am.

If a set of statements could not all be true at the same time, they are said to be **INCONSISTENT**. Otherwise, they are **CONSISTENT**.

We can ask about the consistency of any number of statements. For example, consider the following list of statements:

- (a) There are at least four giraffes at the wild animal park.
- (b) There are exactly seven gorillas at the wild animal park.
- (c) There are not more than two Martians at the wild animal park.
- (d) Every giraffe at the wild animal park is a Martian.

Statements (a) and (d) together imply that there are at least four Martian giraffes at the park. This conflicts with (c), which implies that there are no more than two Martian giraffes there. So the set of statements (a)–(d) is inconsistent. Notice that the inconsistency has nothing at all to do with (b). Statement (b) just happens to be part of an inconsistent set.

Sometimes, people will say that an inconsistent set of statements “contains a contradiction.” By this, they mean that it would be logically impossible for all of the statements to be true at once. A set can be inconsistent even when all of the statements in it are either contingent or tautologous. When a single statement is a contradiction, then that statement alone cannot be true.

### Key Terms

#### 1. Artificial language

2. **Bivalent**
3. **Consistent**
4. **Contingent statement**
5. **Contradiction**
6. **Contradictories**
7. **Formal language**
8. **Formal logic as concern for logical form**
9. **Formal logic as strictly following rules**
10. **Inconsistent**
11. **Logically equivalent**
12. **Natural language**
13. **Tautology**
14. **Truth value**

## **Part II**

# **Critical Thinking**





4

*What is Critical Thinking?*



## 5

### *Informal Fallacies*

#### *What are informal fallacies?*

A fallacy is simply a mistake in reasoning. Some fallacies are formal and some are informal. In chapter ??, we saw that we could define validity formally and thus could determine whether an argument was valid or invalid without even having to know or understand what the argument was about. We saw that we could define certain valid rules of inference, such as *modus ponens* and *modus tollens*. These inference patterns are valid in virtue of their form, not their content. That is, any argument that has the same form as *modus ponens* or *modus tollens* will automatically be valid. A formal fallacy is simply an argument whose form is invalid. Thus, any argument that has that form will automatically be invalid, regardless of the meaning of the sentences. Two formal fallacies that are similar to, but should never be confused with, *modus ponens* and *modus tollens* are denying the antecedent and affirming the consequent. Here are the forms of those invalid inferences:

Denying the antecedent  $P \rightarrow Q \quad \sim P \quad P \rightarrow Q \quad Q \therefore P$   
 $\therefore \quad \sim Q$  Affirming the consequent

Any argument that has either of these forms is an invalid argument. For example:

1. If Kant was a deontologist, then he was a non-consequentialist.
2. Kant was not a deontologist.
3. Therefore, Kant was a not a non-consequentialist.

The form of this argument is:

1.  $D \rightarrow C$
2.  $\sim D$

3.  $\therefore \sim C$

As you can see, this argument has the form of the fallacy, denying the antecedent. Thus, we know that this argument is invalid even if we don't know what "Kant" or "deontologist" or "non-consequentialist" means. ("Kant" was a famous German philosopher from the early 1800s, whereas "deontology" and "non-consequentialist" are terms that come from ethical theory.) It is mark of a formal fallacy that we can identify it even if we don't really understand the meanings of the sentences in the argument. Consider the following argument. It's an argument which uses silly, made-up words from Lewis Carroll's "Jabberwocky." See if you can determine whether the argument's form is valid or invalid:

1. If toves are brillig then toves are slithy.
2. Toves are slithy
3. Therefore, toves are brillig.

You should be able to see that this argument has the form of affirming the consequent:

1.  $B \rightarrow S$
2.  $S$
3.  $\therefore B$

As such, we know that the argument is invalid, even though we haven't got a clue what "toves" are or what "slithy" or "brillig" means. The point is that we can identify formal fallacies without having to know what they mean. In contrast, informal fallacies are those which cannot be identified without understanding the concepts involved in the argument. A paradigm example of an informal fallacy is the fallacy of composition. We will consider this fallacy in the next section. In the remaining sections, we will consider a number of other informal logical fallacies.

### *Fallacy of Composition*

Consider the following argument: Each member on the gymnastics team weighs less than 110 lbs. Therefore, the whole gymnastics team weighs less than 110 lbs.

This arguments commits the composition fallacy. In the composition fallacy one argues that since each part of the whole has a certain feature, it follows that the whole has that same feature. However, you cannot generally identify any argument that moves from statements about parts to statements about wholes as committing the composition fallacy because whether or not there is a fallacy depends on what feature we are attributing

to the parts and wholes. Here is an example of an argument that moves from claims about the parts possessing a feature to a claim about the whole possessing that same feature, but doesn't commit the composition fallacy: Every part of the car is made of plastic. Therefore, the whole car is made of plastic. This conclusion does follow from the premises; there is no fallacy here. The difference between this argument and the preceding argument (about the gymnastics team) isn't their form. In fact both arguments have the same form: Every part of X has the feature f. Therefore, the whole X has the feature f. And yet one of the arguments is clearly fallacious, while the other isn't. The difference between the two arguments is not their form, but their content. That is, the difference is what feature is being attributed to the parts and wholes. Some features (like weighing a certain amount) are such that if they belong to each part, then it does not follow that they belong to the whole. Other features (such as being made of plastic) are such that if they belong to each part, it follows that they belong to the whole. Here is another example: Every member of the team has been to Paris. Therefore the team has been to Paris. The conclusion of this argument does not follow. Just because each member of the team has been to Paris, it doesn't follow that the whole team has been to Paris, since it may not have been the case that each individual was there at the same time and was there in their capacity as a member of the team. Thus, even though it is plausible to say that the team is composed of every member of the team, it doesn't follow that since every member of the team has been to Paris, the whole team has been to Paris. Contrast that example with this one:

Every member of the team was on the plane. Therefore, the whole team was on the plane. This argument, in contrast to the last one, contains no fallacy. It is true that if every member is on the plane then the whole team is on the plane. And yet these two arguments have almost exactly the same form. The only difference is that the first argument is talking about the property, having been to Paris, whereas the second argument is talking about the property, being on the plane. The only reason we are able to identify the first argument as committing the composition fallacy and the second argument as not committing a fallacy is that we understand the relationship between the concepts involved. In the first case, we understand that it is possible that every member could have been to Paris without the team ever having been; in the second case we understand that as long as every member of the team is on the plane, it has to be true that the whole team is on the plane. The take home point here is that in order to identify whether an argument has committed the composition fallacy, one must understand the concepts involved in the argument. This is the mark of an informal fallacy: we have to rely on our understanding of the meanings of the words or concepts involved, rather than simply being able to identify the fallacy from its form.

### *Fallacy of Division*

The division fallacy is like the composition fallacy and they are easy to confuse. The difference is that the division fallacy argues that since the whole has some feature, each part must also have that feature. The composition fallacy, as we have just seen, goes in the opposite direction: since each part has some feature, the whole must have that same feature. Here is an example of a division fallacy: The house costs 1 million dollars. Therefore, each part of the house costs 1 million dollars. This is clearly a fallacy. Just because the whole house costs 1 million dollars, it doesn't follow that each part of the house costs 1 million dollars. However, here is an argument that has the same form, but that doesn't commit the division fallacy:

The whole team died in the plane crash. Therefore each individual on the team died in the plane crash. In this example, since we seem to be referring to one plane crash in which all the members of the team died ("the" plane crash), it follows that if the whole team died in the crash, then every individual on the team died in the crash. So this argument does not commit the division fallacy. In contrast, the following argument has exactly the same form, but does commit the division fallacy: The team played its worst game ever tonight. Therefore, each individual on the team played their worst game ever tonight. It can be true that the whole team played its worst game ever even if it is true that no individual on the team played their worst game ever. Thus, this argument does commit the fallacy of division even though it has the same form as the previous argument, which doesn't commit the fallacy of division. This shows (again) that in order to identify informal fallacies (like composition and division), we must rely on our understanding of the concepts involved in the argument. Some concepts (like "team" and "dying in a plane crash") are such that if they apply to the whole, they also apply to all the parts. Other concepts (like "team" and "worst game played") are such that they can apply to the whole even if they do not apply to all the parts.

### *Begging the question*

Consider the following argument: Capital punishment is justified for crimes such as rape and murder because it is quite legitimate and appropriate for the state to put to death someone who has committed such heinous and inhuman acts. The premise indicator, "because" denotes the premise and (derivatively) the conclusion of this argument. In standard form, the argument is this: 1. It is legitimate and appropriate for the state to put to death someone who commits rape or murder. 2. Therefore, capital punishment is justified for crimes such as rape and murder.

You should notice something peculiar about this argument: the premise is essentially the same claim as the conclusion. The only difference is

that the premise spells out what capital punishment means (the state putting criminals to death) whereas the conclusion just refers to capital punishment by name, and the premise uses terms like “legitimate” and “appropriate” whereas the conclusion uses the related term, “justified.” But these differences don’t add up to any real differences in meaning. Thus, the premise is essentially saying the same thing as the conclusion. This is a problem: we want our premise to provide a reason for accepting the conclusion. But if the premise is the same claim as the conclusion, then it can’t possibly provide a reason for accepting the conclusion! Begging the question occurs when one (either explicitly or implicitly) assumes the truth of the conclusion in one or more of the premises. Begging the question is thus a kind of circular reasoning. One interesting feature of this fallacy is that formally there is nothing wrong with arguments of this form. Here is what I mean. Consider an argument that explicitly commits the fallacy of begging the question. For example, 1. Capital punishment is morally permissible 2. Therefore, capital punishment is morally permissible Now, apply any method of assessing validity to this argument and you will see that it is valid by any method. If we use the informal test (by trying to imagine that the premises are true while the conclusion is false), then the argument passes the test, since any time the premise is true, the conclusion will have to be true as well (since it is the exact same statement). Likewise, the argument is valid by our formal test of validity, truth tables. But while this argument is technically valid, it is still a really bad argument. Why? Because the point of giving an argument in the first place is to provide some reason for thinking the conclusion is true for those who don’t already accept the conclusion. But if one doesn’t already accept the conclusion, then simply restating the conclusion in a different way isn’t going to convince them. Rather, a good argument will provide some reason for accepting the conclusion that is sufficiently independent of that conclusion itself. Begging the question utterly fails to do this and this is why it counts as an informal fallacy. What is interesting about begging the question is that there is absolutely nothing wrong with the argument formally.

Whether or not an argument begs the question is not always an easy matter to sort out. As with all informal fallacies, detecting it requires a careful understanding of the meaning of the statements involved in the argument. Here is an example of an argument where it is not as clear whether there is a fallacy of begging the question: Christian belief is warranted because according to Christianity there exists a being called “the Holy Spirit” which reliably guides Christians towards the truth regarding the central claims of Christianity.

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

One might think that there is a kind of circularity (or begging the question) involved in this argument since the argument appears to assume the truth of Christianity in justifying the claim that Christianity is true. But whether or not this argument really does beg the question is something on which there is much debate within the sub-field of philosophy called epistemology (“study of knowledge”). The philosopher Alvin Plantinga argues persuasively that the argument does not beg the question, but being able to assess that argument takes patient years of study in the field of epistemology (not to mention a careful engagement with Plantinga’s work). As this example illustrates, the issue of whether an argument begs the question requires us to draw on our general knowledge of the world. This is the mark of an informal, rather than formal, fallacy.

### *False dichotomy*

Suppose I were to argue as follows: Raising taxes on the wealthy will either hurt the economy or it will help it. But it won’t help the economy. Therefore it will hurt the economy. The standard form of this argument is:

1. Either raising taxes on the wealthy will hurt the economy or it will help it.
2. Raising taxes on the wealthy won’t help the economy.
3. Therefore, raising taxes on the wealthy will hurt the economy.

This argument contains a fallacy called a “false dichotomy.” A false dichotomy is simply a disjunction that does not exhaust all of the possible options. In this case, the problematic disjunction is the first premise: either raising the taxes on the wealthy will hurt the economy or it will help it. But these aren’t the only options. Another option is that raising taxes on the wealthy will have no effect on the economy. Notice that the argument above has the form of a disjunctive syllogism:

1.  $A \vee B$
2.  $\sim A$
3.  $\therefore B$

However, since the first premise presents two options as if they were the only two options, when in fact they aren’t, the first premise is false and the argument fails. Notice that the form of the argument is perfectly good—the argument is valid. The problem is that this argument isn’t sound because the first premise of the argument commits the false dichotomy fallacy. False dichotomies are commonly encountered in the context of a disjunctive syllogism or constructive dilemma (see chapter 2). In a speech made on April 5, 2004, President Bush made the following remarks about the causes of the Iraq war: Saddam Hussein once again defied the demands of the world. And so I had a choice: Do I take the word of a madman, do I trust a person who had used weapons of mass destruction on his own people, plus



people in the neighborhood, or do I take the steps necessary to defend the country? Given that choice, I will defend America every time. The false dichotomy here is the claim that: Either I trust the word of a madman or I defend America (by going to war against Saddam Hussein's regime). The problem is that these aren't the only options. Other options include ongoing diplomacy and economic sanctions. Thus, even if it true that Bush shouldn't have trusted the word of Hussein, it doesn't follow that the only other

option is going to war against Hussein's regime. (Furthermore, it isn't clear in what sense this was needed to defend America.) That is a false dichotomy. As with all the previous informal fallacies we've considered, the false dichotomy fallacy requires an understanding of the concepts involved. Thus, we have to use our understanding of world in order to assess whether a false dichotomy fallacy is being committed or not.

### *Equivocation*

Consider the following argument: Children are a headache. Aspirin will make headaches go away. Therefore, aspirin will make children go away. This is a silly argument, but it illustrates the fallacy of equivocation. The problem is that the word "headache" is used equivocally—that is, in two different senses. In the first premise, "headache" is used figuratively, whereas in the second premise "headache" is used literally. The argument is only successful if the meaning of "headache" is the same in both premises. But it isn't and this is what makes this argument an instance of the fallacy of equivocation. Here's another example: Taking a logic class helps you learn how to argue. But there is already too much hostility in the world today, and the fewer arguments the better. Therefore, you shouldn't take a logic class. In this example, the word "argue" and "argument" are used equivocally. Hopefully, at this point in the text, you recognize the difference. (If not, go back and reread section 1.1.) The fallacy of equivocation is not always so easy to spot. Here is a trickier example: The existence of laws depends on the existence of intelligent beings like humans who create the laws. However, some laws existed before there were any humans (e.g., laws of physics). Therefore, there must be some non-human, intelligent being that created these laws of nature.

The term "law" is used equivocally here. In the first premise it is used to refer to societal laws, such as criminal law; in the second premise it is used to refer to laws of nature. Although we use the term "law" to apply to both cases, they are importantly different. Societal laws, such as the criminal law of a society, are enforced by people and there are punishments for breaking the laws. Natural laws, such as laws of physics, cannot be broken and thus there are no punishments for breaking them. (Does it make sense to scold the electron for not doing what the law says it will do?) As with every informal fallacy we have examined in this section, equivocation can

only be identified by understanding the meanings of the words involved. In fact, the definition of the fallacy of equivocation refers to this very fact: the same word is being used in two different senses (i.e., with two different meanings). So, unlike formal fallacies, identifying the fallacy of equivocation requires that we draw on our understanding of the meaning of words and of our understanding of the world, generally.

### *Slippery slope fallacies*

Slippery slope fallacies depend on the concept of vagueness. When a concept or claim is vague, it means that we don't know precisely what claim is being made, or what the boundaries of the concept are. The classic example used to illustrate vagueness is the "sorites paradox." The term "sorites" is the Greek term for "heap" and the paradox comes from ancient Greek philosophy. Here is the paradox. I will give you two claims that each sound very plausible, but in fact lead to a paradox. Here are the two claims: 1. One grain of sand is not a heap of sand. 2. If I start with something that is not a heap of sand, then adding one grain of sand to that will not create a heap of sand.

For example, two grains of sand is not a heap, thus (by the second claim) neither is three grains of sand. But since three grains of sand is not a heap then (by the second claim again) neither is four grains of sand. You can probably see where this is going. By continuing to add one grain of sand over and over, I will eventually end up with something that is clearly a heap of sand, but that won't be counted as a heap of sand if we accept both claims 1 and 2 above.

Philosophers continue to argue and debate about how to resolve the sorites paradox, but the point for us is just to illustrate the concept of vagueness. The concept "heap" is a vague concept in this example. But so are so many other concepts, such as color concepts (red, yellow, green, etc.), moral concepts (right, wrong, good, bad), and just about any other concept you can think of. The one domain that seems to be unaffected by vagueness is mathematical and logical concepts. There are two fallacies related to vagueness: the causal slippery slope and the conceptual slippery slope. We'll cover the conceptual slippery slope first since it relates most closely to the concept of vagueness I've explained above.

### *Conceptual slippery slope*

It may be true that there is no essential difference between 499 grains of sand and 500 grains of sand. But even if that is so, it doesn't follow that there is no difference between 1 grain of sand and 5 billion grains of sand. In general, just because we cannot draw a distinction between A and B, and we cannot draw a distinction between B and C, it doesn't mean we cannot

draw a distinction between A and C. Here is an example of a conceptual slippery slope fallacy. It is illegal for anyone under 21 to drink alcohol. But there is no difference between someone who is 21 and someone who is 20 years 11 months old. So there is nothing wrong with someone who is 20 years and 11 months old drinking. But since there is no real distinction between being one month older and one month younger, there shouldn't be anything wrong with drinking at any age. Therefore, there is nothing wrong with allowing a 10 year old to drink alcohol. Imagine the life of an individual in stages of 1 month intervals. Even if it is true that there is no distinction in kind between any one of those stages, it doesn't follow that there isn't a distinction to be drawn at the extremes of either end. Clearly there is a difference between a 5 year old and a 25 year old—a distinction in kind that is relevant to whether they should be allowed to drink alcohol. The conceptual slippery slope fallacy assumes that because we cannot draw a distinction between adjacent stages, we cannot draw a distinction at all between any stages. One clear way of illustrating this is with color. Think of a color spectrum from purple to red to orange to yellow to green to blue. Each color grades into the next without there being any distinguishable boundaries

between the colors—a continuous spectrum. Even if it is true that for any two adjacent hues on the color wheel, we cannot distinguish between the two, it doesn't follow from this that there is no distinction to be drawn between any two portions of the color wheel, because then we'd be committed to saying that there is no distinguishable difference between purple and yellow! The example of the color spectrum illustrates the general point that just because the boundaries between very similar things on a spectrum are vague, it doesn't follow that there are no differences between any two things on that spectrum. Whether or not one will identify an argument as committing a conceptual slippery slope fallacy, depends on the other things one believes about the world. Thus, whether or not a conceptual slippery slope fallacy has been committed will often be a matter of some debate. It will itself be vague. Here is a good example that illustrates this point. People are found not guilty by reason of insanity when they cannot avoid breaking the law. But people who are brought up in certain deprived social circumstances are not much more able than the legally insane to avoid breaking the law. So we should not find such individuals guilty any more than those who are legally insane. Whether there is conceptual slippery slope fallacy here depends on what you think about a host of other things, including individual responsibility, free will, the psychological and social effects of deprived social circumstances such as poverty, lack of opportunity, abuse, etc. Some people may think that there are big differences between those who are legally insane and those who grow up in deprived social circumstances. Others may not think the differences are so great. The issues here are subtle, sensitive, and complex, which is why it is difficult to deter-

mine whether there is any fallacy here or not. If the differences between those who are insane and those who are the product of deprived social circumstances turn out to be like the differences between one shade of yellow and an adjacent shade of yellow, then there is no fallacy here. But if the differences turn out to be analogous to those between yellow and green (i.e., with many distinguishable stages of difference between) then there would indeed be a conceptual slippery slope fallacy here. The difficulty of distinguishing instances of the conceptual slippery slope fallacy, and the fact that distinguishing it requires us to draw on our knowledge about the world, shows that the conceptual slippery slope fallacy is an informal fallacy.

### *Causal slippery slope fallacy*

The causal slippery slope fallacy is committed when one event is said to lead to some other (usually disastrous) event via a chain of intermediary events. If you have ever seen Direct TV's "get rid of cable" commercials, you will know exactly what I'm talking about. (If you don't know what I'm talking about you should Google it right now and find out. They're quite funny.) Here is an example of a causal slippery slope fallacy (it is adapted from one of the Direct TV commercials): If you use cable, your cable will probably go on the fritz. If your cable is on the fritz, you will probably get frustrated. When you get frustrated you will probably hit the table. When you hit the table, your young daughter will probably imitate you. When your daughter imitates you, she will probably get thrown out of school. When she gets thrown out of school, she will probably meet undesirables. When she meets undesirables, she will probably marry undesirables. When she marries undesirables, you will probably have a grandson with a dog collar. Therefore, if you use cable, you will probably have a grandson with dog collar. This example is silly and absurd, yes. But it illustrates the causal slippery slope fallacy. Slippery slope fallacies are always made up of a series of conjunctions of probabilistic conditional statements that link the first event to the last event. A causal slippery slope fallacy is committed when one assumes that just because each individual conditional statement is probable, the conditional that links the first event to the last event is also probable. Even if we grant that each "link" in the chain is individually probable, it doesn't follow that the whole chain (or the conditional that links the first event to the last event) is probable. Suppose, for the sake of the argument, we assign probabilities to each "link" or conditional statement, like this. (I have italicized the consequents of the conditionals and assigned high conditional probabilities to them. The high probability is for the sake of the argument; I don't actually think these things are as probable as I've assumed here.)

- If you use cable, then your cable will probably go on the fritz (.9)

- If your cable is on the fritz, then you will probably get angry (.9)
- If you get angry, then you will probably hit the table (.9)
- If you hit the table, your daughter will probably imitate you (.8)
- If your daughter imitates you, she will probably be kicked out of school (.8)
- If she is kicked out of school, she will probably meet undesirables (.9)
- If she meets undesirables, she will probably marry undesirables (.8)
- If she marries undesirables, you will probably have a grandson with a dog collar (.8)

However, even if we grant the probabilities of each link in the chain is high (80- 90% probable), the conclusion doesn't even reach a probability higher than chance. Recall that in order to figure the probability of a conjunction, we must multiply the probability of each conjunct:  $(.9) \times (.9) \times (.9) \times (.8) \times (.8) \times (.9) \times (.8) \times (.8) = .27$  That means the probability of the conclusion (i.e., that if you use cable, you will have a grandson with a dog collar) is only 27%, despite the fact that each conditional has a relatively high probability! The causal slippery slope fallacy is actually a formal probabilistic fallacy and so could have been discussed in chapter 3 with the other formal probabilistic fallacies. What makes it a formal rather than informal fallacy is that we can identify it without even having to know what the sentences of the argument mean. I could just have easily written out a nonsense argument comprised of series of probabilistic conditional statements. But I would still have been able to identify the causal slippery slope fallacy because I would have seen that there was a series of probabilistic conditional statements leading to a claim that the conclusion of the series was also probable. That is enough to tell me that there is a causal slippery slope fallacy, even if I don't really understand the meanings of the conditional statements. It is helpful to contrast the causal slippery slope fallacy with the valid form of inference, hypothetical syllogism. Recall that a hypothetical syllogism has the following kind of form:

- $A \rightarrow B$
- $B \rightarrow C$
- $C \rightarrow D$
- $D \rightarrow E$
- $\therefore A \rightarrow E$

The only difference between this and the causal slippery slope fallacy is that whereas in the hypothetical syllogism, the link between each component is certain, in a causal slippery slope fallacy, the link between each event is probabilistic. It is the fact that each link is probabilistic that accounts for the fallacy. One way of putting this is point is that probability is not transitive. Just because A makes B probable and B makes C probable and C makes X probable, it doesn't follow that A makes X probable. In contrast, when the links are certain rather than probable, then if A always leads to B and B always leads to C and C always leads to X, then it has to be the case that A always leads to X.

### *Fallacies of relevance*

What all fallacies of relevance have in common is that they make an argument or response to an argument that is irrelevant. Fallacies of relevance can be compelling psychologically, but it is important to distinguish between rhetorical techniques that are psychologically compelling, on the one hand, and rationally compelling arguments, on the other. What makes something a fallacy is that it fails to be rationally compelling, once we have carefully considered it. That said, arguments that fail to be rationally compelling may still be psychologically or emotionally compelling. The first fallacy of relevance that we will consider, the ad hominem fallacy, is an excellent example a fallacy that can be psychologically compelling.

### *Ad hominem*

“Ad hominem” is a Latin phrase that can be translated into English as the phrase, “against the man.” In an ad hominem fallacy, instead of responding to (or attacking) the argument a person has made, one attacks the person him or herself. In short, one attacks the person making the argument rather than the argument itself. Here is an anecdote that reveals an ad hominem fallacy (and that has actually occurred in my ethics class before). A philosopher named Peter Singer had made an argument that it is morally wrong to spend money on luxuries for oneself rather than give all of your money that you don't strictly need away to charity. The argument is actually an argument from analogy (whose details I discussed in section 3.3), but the essence of the argument is that there are every day in this world children who die preventable deaths, and there are charities who

could save the lives of these children if they are funded by individuals from wealthy countries like our own. Since there are things that we all regularly buy that we don't need (e.g., Starbuck's lattes, beer, movie tickets, or extra clothes or shoes we don't really need), if we continue to purchase those things rather than using that money to save the lives of children, then we are essentially contributing to the deaths of those children if we choose

to continue to live our lifestyle of buying things we don't need, rather than donating the money to a charity that will save lives of children in need. In response to Singer's argument, one student in the class asked: "Does Peter Singer give his money to charity? Does he do what he says we are all morally required to do?" The implication of this student's question (which I confirmed by following up with her) was that if Peter Singer himself doesn't donate all his extra money to charities, then his argument isn't any good and can be dismissed. But that would be to commit an *ad hominem* fallacy. Instead of responding to the argument that Singer had made, this student attacked Singer himself. That is, they wanted to know how Singer lived and whether he was a hypocrite or not. Was he the kind of person who would tell us all that we had to live a certain way but fail to live that way himself? But all of this is irrelevant to assessing Singer's argument. Suppose that Singer didn't donate his excess money to charity and instead spent it on luxurious things for himself. Still, the argument that Singer has given can be assessed on its own merits. Even if it were true that Peter Singer was a total hypocrite, his argument may nevertheless be rationally compelling. And it is the quality of the argument that we are interested in, not Peter Singer's personal life and whether or not he is hypocritical. Whether Singer is or isn't a hypocrite, is irrelevant to whether the argument he has put forward is strong or weak, valid or invalid. The argument stands on its own and it is that argument rather than Peter Singer himself that we need to assess. Nonetheless, there is something psychologically compelling about the question: Does Peter Singer practice what he preaches? I think what makes this question seem compelling is that humans are very interested in finding "cheaters" or hypocrites—those who say one thing and then do another. Evolutionarily, our concern with cheaters makes sense because cheaters can't be trusted and it is essential for us (as a group) to be able to pick out those who can't be trusted. That said, whether or not a person giving an argument is a hypocrite is irrelevant to whether that person's argument is good or bad. So there may be psychological reasons why humans are prone to find certain kinds of *ad*

*hominem* fallacies psychologically compelling, even though *ad hominem* fallacies are not rationally compelling. Not every instance in which someone attacks a person's character is an *ad hominem* fallacy. Suppose a witness is on the stand testifying against a defendant in a court of law. When the witness is cross examined by the defense lawyer, the defense lawyer tries to go for the witness's credibility, perhaps by digging up things about the witness's past. For example, the defense lawyer may find out that the witness cheated on her taxes five years ago or that the witness failed to pay her parking tickets. The reason this isn't an *ad hominem* fallacy is that in this case the lawyer is trying to establish whether what the witness is saying is true or false and in order to determine that we have to know whether the witness is trustworthy. These facts about the witness's past

may be relevant to determining whether we can trust the witness's word. In this case, the witness is making claims that are either true or false rather than giving an argument. In contrast, when we are assessing someone's argument, the argument stands on its own in a way the witness's testimony doesn't. In assessing an argument, we want to know whether the argument is strong or weak and we can evaluate the argument using the logical techniques surveyed in this text. In contrast, when a witness is giving testimony, they aren't trying to argue anything. Rather, they are simply making a claim about what did or didn't happen. So although it may seem that a lawyer is committing an *ad hominem* fallacy in bringing up things about the witness's past, these things are actually relevant to establishing the witness's credibility. In contrast, when considering an argument that has been given, we don't have to establish the arguer's credibility because we can assess the argument they have given on its own merits. The arguer's personal life is irrelevant.

### *Straw Man*

Suppose that my opponent has argued for a position, call it position A, and in response to his argument, I give a rationally compelling argument against position B, which is related to position A, but is much less plausible (and thus much easier to refute). What I have just done is attacked a straw man—a position that “looks like” the target position, but is actually not that position. When one attacks a straw man, one commits the straw man fallacy. The straw man fallacy misrepresents one's opponent's argument and is thus a kind of irrelevance. Here is an example.

Two candidates for political office in Colorado, Tom and Fred, are having an exchange in a debate in which Tom has laid out his plan for putting more money into health care and education and Fred has laid out his plan which includes earmarking more state money for building more prisons which will create more jobs and, thus, strengthen Colorado's economy. Fred responds to Tom's argument that we need to increase funding to health care and education as follows: “I am surprised, Tom, that you are willing to put our state's economic future at risk by sinking money into these programs that do not help to create jobs. You see, folks, Tom's plan will risk sending our economy into a tailspin, risking harm to thousands of Coloradans. On the other hand, my plan supports a healthy and strong Colorado and would never bet our state's economic security on idealistic notions that simply don't work when the rubber meets the road.” Fred has committed the straw man fallacy. Just because Tom wants to increase funding to health care and education does not mean he does not want to help the economy. Furthermore, increasing funding to health care and education does not entail that fewer jobs will be created. Fred has attacked a position that is not the position that Tom holds, but is in fact a much less plausible, easier



to refute position. However, it would be silly for any political candidate to run on a platform that included “harming the economy.” Presumably no political candidate would run on such a platform. Nonetheless, this exact kind of straw man is ubiquitous in political discourse in our country. Here is another example. Nancy has just argued that we should provide middle schoolers with sex education classes, including how to use contraceptives so that they can practice safe sex should they end up in the situation where they are having sex. Fran responds: “proponents of sex education try to encourage our children to a sex-with-no-strings-attached mentality, which is harmful to our children and to our society.” Fran has committed the straw man (or straw woman) fallacy by misrepresenting Nancy’s position. Nancy’s position is not that we should encourage children to have sex, but that we should make sure that they are fully informed about sex so that if they do have sex, they go into it at least a little less blindly and are able to make better decision regarding sex.

As with other fallacies of relevance, straw man fallacies can be compelling on some level, even though they are irrelevant. It may be that part of the reason we are taken in by straw man fallacies is that humans are prone to “demonize” the “other”—including those who hold a moral or political position different from our own. It is easy to think bad things about those with whom we do not regularly interact. And it is easy to forget that people who are different than us are still people just like us in all the important respects. Many years ago, atheists were commonly thought of as highly immoral people and stories about the horrible things that atheists did in secret circulated widely. People believed that these strange “others” were capable of the most horrible savagery. After all, they may have reasoned, if you don’t believe there is a God holding us accountable, why be moral? The Jewish philosopher, Baruch Spinoza, was an atheist who lived in the Netherlands in the 17th century. He was accused of all sorts of things that were commonly believed about atheists. But he was in fact as upstanding and moral as any person you could imagine. The people who knew Spinoza knew better, but how could so many people be so wrong about Spinoza? I suspect that part of the reason is that since at that time there were very few atheists (or at least very few people actually admitted to it), very few people ever knowingly encountered an atheist. Because of this, the stories about atheists could proliferate without being put in check by the facts. I suspect the same kind of phenomenon explains why certain kinds of straw man fallacies proliferate. If you are a conservative and mostly only interact with other conservatives, you might be prone to holding lots of false beliefs about liberals. And so maybe you are less prone to notice straw man fallacies targeted at liberals because the false beliefs you hold about them incline you to see the straw man fallacies as true.

*Tu quoque*

“Tu quoque” is a Latin phrase that can be translated into English as “you too” or “you, also.” The tu quoque fallacy is a way of avoiding answering a criticism by bringing up a criticism of your opponent rather than answer the criticism. For example, suppose that two political candidates, A and B, are discussing their policies and A brings up a criticism of B’s policy. In response, B brings up her own criticism of A’s policy rather than respond to A’s criticism of her policy. B has here committed the tu quoque fallacy. The fallacy is best understood as a way of avoiding having to answer a tough criticism that one may not have a good answer to. This kind of thing happens all the time in political discourse.

Tu quoque, as I have presented it, is fallacious when the criticism one raises is simply in order to avoid having to answer a difficult objection to one’s argument or view. However, there are circumstances in which a tu quoque kind of response is not fallacious. If the criticism that A brings toward B is a criticism that equally applies not only to A’s position but to any position, then B is right to point this fact out. For example, suppose that A criticizes B for taking money from special interest groups. In this case, B would be totally right (and there would be no tu quoque fallacy committed) to respond that not only does A take money from special interest groups, but every political candidate running for office does. That is just a fact of life in American politics today. So A really has no criticism at all to B since everyone does what B is doing and it is in many ways unavoidable. Thus, B could (and should) respond with a “you too” rebuttal and in this case that rebuttal is not a tu quoque fallacy.

*Genetic fallacy*

The genetic fallacy occurs when one argues (or, more commonly, implies) that the origin of something (e.g., a theory, idea, policy, etc.) is a reason for rejecting (or accepting) it. For example, suppose that Jack is arguing that we should allow physician assisted suicide and Jill responds that that idea first was used in Nazi Germany. Jill has just committed a genetic fallacy because she is implying that because the idea is associated with Nazi Germany, there must be something wrong with the idea itself. What she should have done instead is explain what, exactly, is wrong with the idea rather than simply assuming that there must be something wrong with it since it has a negative origin. The origin of an idea has nothing inherently to do with its truth or plausibility. Suppose that Hitler constructed a mathematical proof in his early adulthood (he didn’t, but just suppose). The validity of that mathematical proof stands on its own; the fact that Hitler was a horrible person has nothing to do with whether the proof is good. Likewise with any other idea: ideas must be assessed on their own merits and the origin of an idea is neither a merit nor demerit of the idea. Although genetic fallacies

are most often committed when one associates an idea with a negative origin, it can also go the other way: one can imply that because the idea has a positive origin, the idea must be true or more plausible. For example, suppose that Jill argues that the Golden Rule is a good way to live one's life because the Golden Rule originated with Jesus in the Sermon on the Mount (it didn't, actually, even though Jesus does state a version of the Golden Rule). Jill has committed the genetic fallacy in assuming that the (presumed)

fact that Jesus is the origin of the Golden Rule has anything to do with whether the Golden Rule is a good idea. I'll end with an example from William James's seminal work, *The Varieties of Religious Experience*. In that book (originally a set of lectures), James considers the idea that if religious experiences could be explained in terms of neurological causes, then the legitimacy of the religious experience is undermined. James, being a materialist who thinks that all mental states are physical states—ultimately a matter of complex brain chemistry, says that the fact that any religious experience has a physical cause does not undermine that veracity of that experience. Although he doesn't use the term explicitly, James claims that the claim that the physical origin of some experience undermines the veracity of that experience is a genetic fallacy. Origin is irrelevant for assessing the veracity of an experience, James thinks. In fact, he thinks that religious dogmatists who take the origin of the Bible to be the word of God are making exactly the same mistake as those who think that a physical explanation of a religious experience would undermine its veracity. We must assess ideas for their merits, James thinks, not their origins.

### *Appeal to consequences*

The appeal to consequences fallacy is like the reverse of the genetic fallacy: whereas the genetic fallacy consists in the mistake of trying to assess the truth or reasonableness of an idea based on the origin of the idea, the appeal to consequences fallacy consists in the mistake of trying to assess the truth or reasonableness of an idea based on the (typically negative) consequences of accepting that idea. For example, suppose that the results of a study revealed that there are IQ differences between different races (this is a fictitious example, there is no such study that I know of). In debating the results of this study, one researcher claims that if we were to accept these results, it would lead to increased racism in our society, which is not tolerable. Therefore, these results must not be right since if they were accepted, it would lead to increased racism. The researcher who responded in this way has committed the appeal to consequences fallacy. Again, we must assess the study on its own merits. If there is something wrong with the study, some flaw in its design, for example, then that would be a relevant criticism of the study. However, the fact that the results of the study, if widely circulated, would have a negative effect on society is not a

reason for rejecting these results as false. The consequences of some idea (good or bad) are irrelevant to the truth or reasonableness of that idea.

Notice that the researchers, being convinced of the negative consequences of the study on society, might rationally choose not to publish the study (for fear of the negative consequences). This is totally fine and is not a fallacy. The fallacy consists not in choosing not to publish something that could have adverse consequences, but in claiming that the results themselves are undermined by the negative consequences they could have. The fact is, sometimes truth can have negative consequences and falsehoods can have positive consequences. This just goes to show that the consequences of an idea are irrelevant to the truth or reasonableness of an idea.

### *Appeal to authority*

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

For example, suppose that a group of doctors sign a petition to prohibit abortions, claiming that abortions are morally wrong. If Bob cites that fact that these doctors are against abortion, therefore abortion must be morally wrong, then Bob has committed the appeal to authority fallacy. The problem is that doctors are not authorities on what is morally right or wrong. Even if they are authorities on how the body works and how to perform certain procedures (such as abortion), it doesn't follow that they are authorities on whether or not these procedures ought to be performed—they are not experts on the ethical status of these procedures. It would be just as much an appeal to consequences fallacy if Melissa were to argue that since some other group of doctors supported abortion, that shows that it must be morally acceptable. In either case, since doctors are not authorities on moral issues, their opinions on a moral issue like abortion is irrelevant. In general, an appeal to authority fallacy occurs when someone takes what an individual says as evidence for some claim, when that individual has no particular expertise in the relevant domain (even if they do have expertise in some other, unrelated, domain).

## 6

# Cognitive Biases

### Introduction

Humans are complicated. We have limited access to the world around us and yet we've built elaborate tools to let us access even more than we used to imagine could've existed. Each of us comes equipped with a roughly 1.4 kg hunk of fatty tissue we call the 'brain' composed of approximately 170 billion cells. Roughly half of these are neurons that compose an elaborate electrical network that reacts to and interacts with the world. This network has evolved over several million years in human beings as we've interacted with and altered the world around us.

Unfortunately, the adaptations that made humans well-suited to their environments during the previous three million years or so aren't necessarily the ones that make us perfect logical reasoners. One reason for this is that perfect logical reasoning is *hard work*. Sometimes it's more cost effective to risk being wrong or making a mistake if the alternative requires you to write out a long, formal argument. Another reason is that sometimes we face problems that can't be solved by detailed logical analysis. If you want your crush to reciprocate your feelings towards them, consider that writing them a lengthy proof may not be the best strategy. If you want to resolve a dispute between two aggrieved parties, you may need to do other things before they will listen to a well-reasoned compromise.

Humans are social creatures and so many of the cognitive shortcuts we evolved may be for socially expedient reasons. After all, if you disagree with everyone in your community you are more likely to be made an outcast than to be hailed as a wise seeker of the truth. Socrates was put to death by his fellow Athenians for "worshipping false gods and corrupting the youth."

In this chapter we are going to explore some of the cognitive biases that affect us as we try to reason about the world. Many of our cognitive biases are active objects of study by philosophers, cognitive scientists, and psychologists. Before we begin examining these biases in more detail, there are a few things to remark about cognitive biases in general.

First, and perhaps most important, is that *everyone* experiences these biases from time to time. No one is immune to them, because we all get our basic cognitive structure from the same place. Knowing about these biases gives us an advantage as critical thinkers because we can identify and study these biases, but just knowing about them isn't enough to make us better at avoiding them. To get better at that takes practice and conscious effort. Even then, your brain may revert to using a bias without you even realizing it!

Second, just because these patterns of reasoning are biased doesn't necessarily mean that the conclusions drawn from them are always wrong. We can reason in a biased way and still get to the right conclusion. The problem with biases of the sort we're discussing here is that these patterns of reasoning aren't *reliable* at getting us to the truth. They often lead us astray if another conclusion is more expedient. We want to avoid biased reasoning not because it's always wrong but because it is less consistently right.

Third, we are focusing here on cognitive biases, or biases in our reasoning that effect how human beings actually make inferences. There are other kinds of bias—for example, a bias towards preferring certain kinds of food or music—that aren't classified as cognitive in this way. Likewise, there are biases that are distinctly *moral* in character, such as being biased against people who are different from you. These aren't the kinds of biases we have in mind here.

Finally, this isn't an exhaustive or comprehensive list of cognitive biases. Because this area is being actively studied scientists often discover that we are more or less biased than we thought, or that our biases are more sensitive to environmental or social conditions than we first realized. As a good critical thinker, it's worth your time and effort to stay updated on research into human biases and to develop a habit of reflecting on your own reasoning. You may identify biases in your own reasoning that we don't discuss here!

## Confirmation Bias

**CONFIRMATION BIAS** is the unconscious tendency to discount, ignore, or forget information that contradicts our previously held beliefs. This term was first coined by the psychologist Peter Wason based on his 1960 paper, "On the failure to eliminate hypotheses in a conceptual task."<sup>1</sup> In the paper Wason conducts the following very elegant experiment.

He asks participants to perform what is called the "2-4-6 Task." The participants are told that the experimenter has in mind a rule to identify triples of numbers. Every triple either satisfies the rule or doesn't. Participants are then told that the triple (2, 4, 6) satisfies the rule and they are invited to guess new triples. The experimenter will tell the participants whether or

<sup>1</sup> See .

not their guess satisfies the rule.

**Take a second and think about what the rule Wason used might be. Do you have a guess? What triples would you ask the experimenter for evidence?**

What Wason found was that participants are overwhelmingly likely to guess triples that are *consistent* with their initial guesses about what the rule might be. Most participants guessed extremely specific rules like “The number in the middle is the average of the other two.” or “Three consecutive even numbers.” These participants would suggest triples like (5, 10, 15) or (10, 12, 14), which were confirmed by the experimenter to follow the rule. In fact, the rule is simply “Any three numbers in ascending order.” so triples like (1, 7, 100) or (−2, 0.1, 5) are also admissible.

The point that Wason drew from his experiment is that participants were more likely to seek evidence that *confirmed* their belief about the rule rather than evidence that would *falsify* or contradict that belief.

How do we fall for confirmation bias in everyday life? One common way is to surround ourselves with sources of evidence—including friends, news sources, or podcasts—that *confirm* our beliefs rather than challenge them. Social media and technology companies use algorithms to provide us with more of what we like. Our horoscope tells us something about ourselves that we already believe, leading us to trust the horoscope.

This might not be so bad if we are looking for movie or restaurant recommendations, but it can have deleterious effects on other aspects of our life. If we are looking for advice for school or our love lives, we might listen to people telling us things we already believe rather than challenging us. Someone who believes they are entitled to something will listen to the person who confirms that belief. Instead, we can fight back against confirmation bias by assessing our sources of evidence impartially, by seeking out multiple sources of evidence, and asking yourself what it would take to change your mind.

This last technique can be especially effective if applied in advance. Consider a person who believes that the Earth is flat. Now suppose this person asked themselves, “What would it take for me to believe that the Earth is not flat, but is actually round?” They might decide that *absolutely nothing* could ever convince them otherwise. This is a good sign of bad faith reasoning. Most of our beliefs should be sensitive to *some kind of evidence*, even if that evidence would be hard to obtain. Suppose the flat-Earth believer decides to conduct an experiment for themselves, and they decide if the experimental results go one way they will change their mind and if the results go another way they will stick to their position.

This is the kind of thing that you should do regularly! Consider what sorts of evidence would change your mind. By doing so in advance of collecting evidence you avoid a common fallacy called **MOVING THE GOAL-POSTS**. This fallacy occurs when some evidence is considered sufficient for

some claim  $P$ , only to deny  $P$  once the evidence is provided.

### *Availability Heuristic*

The **AVAILABILITY HEURISTIC** (also called the “availability bias”) is closely related to confirmation bias. This bias occurs when we only attend to the information that is readily available to us. Sometimes this is referred to as the “selective attention bias.” The idea is that human beings have a limited cognitive capacity and so we’re more likely to judge as important the things that we can easily remember or observe, and less likely to judge as important things that are harder to remember or observe.

Psychological work by Kahneman and Tversky showed that people are more likely to remember things that are more “salient” in some way.<sup>2</sup> For example, if ask to memorize a list of names, some of which are ordinary and others are the names of celebrities, participants remember the celebrity names better than the ordinary ones.

Try this. Which do you think is more common: English words that begin with the letter K or English words that have K as the third letter? You probably have an easier time coming up with examples of the former over the latter.

As it turns out, there are roughly three times as many words with K in the third letter position as there are words that begin with K! The explanation for our tendency to think otherwise is that we tend to emphasize the evidence we can recall and deemphasize what we can’t. This may be unsurprising, given that we can’t treat as evidence something we can’t even remember! But it is important to recognize that we may not be seeing all the facts, even if we are trying to collect evidence. Especially when generalizing about something, pay attention to where you are getting your evidence from. Did you make an attempt to collect evidence in an unbiased way?

To avoid falling prey to bad reasoning because of the availability heuristic, try to ask yourself what sources of evidence you may be missing, and to pay special attention to disconfirming evidence.

### *Anchoring and Ordering Effects*

A psychological consequence of availability is that we often pay attention to only *part* of the available evidence. There are many ways that this can occur; so many so that it would be too large a task to try and list them all.

However, we can categorize many forms of availability bias into **ANCHORING** and **ORDERING EFFECTS**.

Anchoring occurs when we receive some collection of evidence and we use that evidence to establish an unjustified ‘norm’ or *anchor* against which we compare future evidence. A very common form of anchoring is called



‘price anchoring’ and it is well known to advertisers and corporations with an interest in selling you things.

Here’s an example of price anchoring. Suppose an ice-cream shop sells three sizes of ice cream: an 6 oz. cup for \$2, a 16 oz cup for \$6, and a 48 oz bucket for \$20. Many people will see the \$20 bucket of ice cream and ‘anchor’ their expectations about the price of the ice cream based on that price. Few people may purchase the \$20 bucket, but more people may be inclined to get the \$6 cup over the \$2 because it’s the “middle choice.”

Consider another example of price anchoring: sale pricing. By telling customers that a \$50 item is now on sale for \$30, the customer will be more likely to believe that the product is *worth* \$50 or that that is the fair price. This may or may not be the case, but more often it makes customers more likely to buy an expensive item if they believe that the true value is even higher. Sometimes, price anchoring can even have the straightforward effect of causing others to believe that a more expensive item is higher quality. This is known as the pricing bias.

Anchoring doesn’t only occur in pricing. We also anchor our expectations based on the order that we observe things. **ORDERING EFFECTS** occur when the order in which we receive information effects our evaluation and recall of that information. For example, if we are given evidence in a sequence and are more likely to recall the first or last pieces of evidence.

Sometimes the order of information can effect our decisions or evaluations as well. In another study by Kahneman and Tversky, participants were told they would have to quickly calculate a long mathematical equation. Participants were then shown either  $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8 = ?$  or  $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = ?$ .

Because the time that participants were given was so short, they effectively had to reply with an estimate of the answer rather than an explicit calculation. When participants were shown the former equation (beginning with smaller numbers) the median estimate was 512; when they were shown the latter equation (beginning with larger numbers), the median estimate was 2,250. In both cases the participants dramatically underestimated the value, but the underestimation was even worse for participants who began reading the equation with smaller numbers.<sup>3</sup>

<sup>3</sup> The correct answer is orders of magnitude larger than the median guesses: 40,320.

### *Survivorship Bias*

During World War II, the U.S. government commissioned many academics to support the war effort. Among them were statisticians at the Statistical Research Group (SRG) at Columbia University, including Abraham Wald. One of the projects that Wald contributed to was identifying how to protect bombers from being shot down.

Consider the plane depicted in figure 6.1. Military officials at the Center for Naval Analyses had identified the areas where bombers tended to

exhibit damage (marked as red dots in the diagram) and were designing ways to add armor to those areas that exhibited the most damage (the wings, center of the fuselage, and the tail). It was thought that by armoring the areas of the planes that showed the most damage the planes would be more likely to survive enemy attacks.

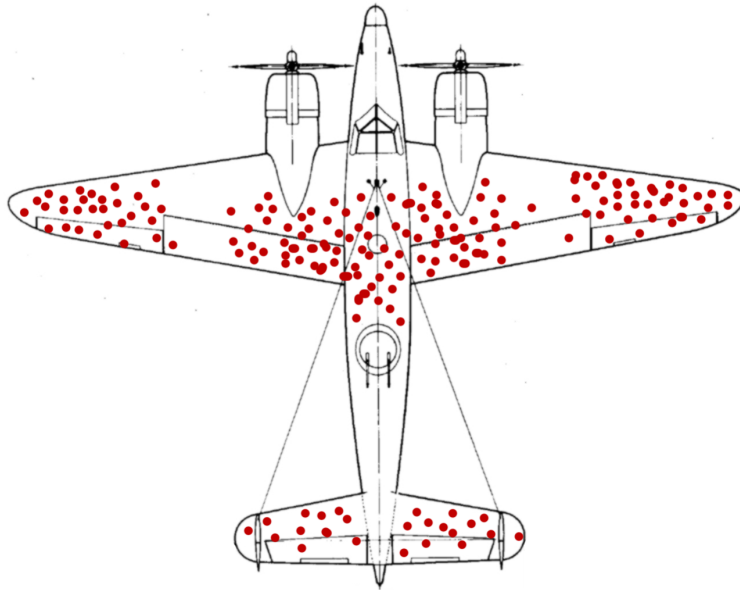


Figure 6.1: A diagram of the locations where damage was recorded on bombers during WWII. Source: <https://commons.wikimedia.org/wiki/File:Survivorship-bias.png>

Wald observed in response to this plan that the data was only collected from the bombers that had *safely returned* from their bombing runs. That is, the damage these planes exhibited was damage that the planes could withstand and still return home. Instead, the military should armor the *undamaged* parts of the plane. This resulted in far fewer planes being shot down and more pilots able to return home.

This is an example of the **SURVIVORSHIP BIAS**, or a bias towards favoring or focusing on data collected from sources that *survived* some filter or process, rather than being eliminated by it. Prominent examples of survivorship bias include the following:

1. The belief that the only abortions that occur are the ones reported in hospital records (many abortions occur at home in unsafe conditions due to poverty and lack of healthcare).
2. The belief that music, television, or other media used to be better (there used to be lots of terrible media too, but we don't listen to it any more).
3. The belief that successful CEOs or entrepreneurs have some special ability or talent that enables them to succeed (there are many people

who fail due to pure randomness or lack of resources and successful people often start out with many more advantages).

Can you think of other examples of conclusions drawn from evidence biased in this way?

### *Fundamental Attribution Error*

We don't have direct access to anyone's intentions except our own. When we act, we know (more or less) *why* we acted and what we had hoped to achieve by acting. Compare that to how much we know about why other people act. Most of the time we have very little access to anyone else's intentions. We can't hear their thoughts. Sometimes people tell us what they want to achieve, other times we make inferences about what we think their intentions were. This is a difficult task and unsurprisingly we are biased in various ways when we try to achieve it.

The **FUNDAMENTAL ATTRIBUTION ERROR** is the phenomenon of explaining our own behavior mostly in terms of *external* influences like our environment, our circumstances, an illness or a new stressor, etc. while explaining the behavior of *others* in terms of internal characteristics like their character, beliefs, or desires.

As George Carlin once remarked, "Have you ever noticed that anybody driving slower than you is an idiot, and anyone going faster than you is a maniac?" The fundamental attribution error causes us to explain away our own behavior and blame others for theirs.

### *Emotive Biases*

Many philosophers once believed that our emotions and our faculty of reason were in direct competition. They thought that rationality was somehow impeded or stymied by the emotions and that it would be better if we could remain dispassionate, unemotional, and detached. We now know better. Reason and emotion are woven together in a complicated and largely beneficial way. We can measure the ways in which emotion and reason interact both positively and negatively and identify circumstances where emotion can be biasing or completely neutral.

In general, emotive biases are a kind of in which some subject, image, or sensation elicits an emotional response. For example, a strong positive emotional response can sometimes be associated with increased risk taking.<sup>4</sup> Here are a few other examples of affect biases:

**Negativity Bias:** A negativity bias occurs when we attach more evidential significance to negative experiences than to positive ones. This is often manifest in a belief that negative information is more likely to be true than positive information. Consider political advertisements that attack an op-

<sup>4</sup> See e.g. Isen and Patrick (1983) and Anderson and Galinsky (2006).

posing candidate, rather than ads that promote the benefits of a candidate's policies.

**Euphemism/Dysphemism:** Euphemisms and dysphemisms are alternative terms for concepts that are either generally thought of as bad or good, respectively. In other words, these terms are used to shift the *connotation* of a term to have a different value and to elicit a different emotional reaction. An example of a euphemism would be to substitute the term “freedom fighter” in place of the term “terrorist.” “Terrorist” has a generally negative emotional connotation, while the term “freedom fighter” seems more positive.

A dysphemism has a similar effect but in the reverse direction. A dysphemism takes a word or phrase with a generally positive or neutral connotation and replaces it with one that connotes something more negative. The intent is often to communicate that something or someone is bad or immoral by using a more grotesque or dehumanizing term for that person or thing. Slurs and insults can be dysphemisms and dysphemisms can also vary cross-culturally or in different conversational contexts.

Since different concepts elicit different emotional responses in people, what counts as a euphemism or dysphemism is not typically universal. However, it is important to recognize the conversational context one is in or the effects that certain types of communication can have. Context matters, and what once counted as a euphemism may eventually take on the negative affect of the concept it was applied to. Pay attention to the affective response that words have on others to avoid biasing your arguments.

### *Zeigarnik Effect*

Bluma Zeigarnik was a Soviet psychologist who, among other things, studied memory. After observing that waiters in a cafe were better at remembering the contents of someone's order before they had paid, but after paying the waiter essentially forgot the entire order, Zeigarnik designed several experiments to test this phenomenon.<sup>5</sup> The effect she found has come to be known as the **ZEIGARNIK EFFECT**

5

What she found was that people tend to remember things more accurately and readily if they were interrupted in the middle of their task when compared to cases where they were allowed to complete the task. The proposed explanation is that our brain discards memories that are no longer relevant to us once the task is complete, but uncompleted tasks continue to impose a cognitive load.

### *Key Terms*

1. **Affect Bias**
2. **Anchoring**
3. **Cognitive Bias**
4. **Confirmation Bias**
5. **Dysphemism**
6. **Euphemism**
7. **Fundamental Attribution Error**
8. **Moving the Goalposts**
9. **Negativity**
10. **Ordering Effects**
11. **Survivorship Bias**
12. **Zeigarnik Effect**



## **Part III**

# **Categorical Logic**





## 7

# Categorical Statements

### *Quantified Categorical Statements*

Back in Chapter 1, we saw that a statement was a unit of language that could be true or false (a proposition is the “content” of a statement). In this chapter and the next we are going to look at a particular kind of statement, called a *quantified categorical* statement, and begin to develop a formal theory of how to create arguments using these statements. This kind of logic is generally called “categorical” or “Aristotelian” logic, because it was originally invented by the great logician and philosopher Aristotle in the fourth century BCE. This kind of logic was the preeminent logical formalism throughout Europe, Africa, and the Middle East for 20 centuries afterward, and was expanded in all kinds of fascinating ways by Muslim and Catholic scholars, some of which we will look at here.

Consider the following propositions:

- (a) All dogs are mammals.
- (b) Most physicists are smart.
- (c) Few politicians are honest.
- (d) No dogs are cats.
- (e) Some Americans are doctors.
- (f) Some adults are not logicians.
- (g) Thirty percent of Canadians speak French.
- (h) One chair is missing.

These are all examples of quantified categorical statements. A **QUANTIFIED CATEGORICAL STATEMENT** is a statement that makes a claim about a certain quantity of the members of a class or group. (Sometimes we will just call these “categorical statements”) Statement (a), for example, is about the class of dogs and the class of mammals. These statements make no mention of

any particular members of the categories or classes or types they are about. The propositions are also *quantified* in that they state *how many* of the things in one class are also members of the other. For instance, statement (b) talks about *most* physicists, while statement (c) talks about *few* teachers.

Categorical statements can be broken down into four parts: the quantifier, the subject term, the predicate term, and the copula. The **QUANTIFIER** is the part of a categorical sentence that specifies a portion of a class. It is the “how many” term. The quantifiers in the sentences above are all, most, few, no, some, thirty percent, and one. Notice that the “no” in sentence (d) counts as a quantifier, the same way zero counts as a number. The subject and predicate terms are the two classes the statement talks about. The **SUBJECT CLASS** is the first class mentioned in a quantified categorical statement, and the **PREDICATE CLASS** is the second. In sentence (e), for instance, the subject class is the class of Americans and the predicate class is the class of doctors. The **COPULA** is simply the form of the verb “to be” that links subject and predicate. Notice that the quantifier is always referring to the subject. The statement “Thirty percent of Canadians speak French” is saying something about a portion of Canadians, not about a portion of French speakers.

Sentence (g) is a little different than the others. In sentence (g) the subject is the class of Canadians and the predicate is the class of people who speak French. That’s not quite the way it is written, however. There is no explicit copula, and instead of giving a noun phrase for the predicate term, like “people who speak French,” it has a verb phrase, “speak French.” If you are asked to identify the copula and predicate for a sentence like this, you should say that the copula is implicit and transform the verb phrase into a noun phrase. You would do something similar for sentence (h): the subject term is “chair,” and the predicate term is “things that are missing.” We will go into more detail about these issues in Section 7.

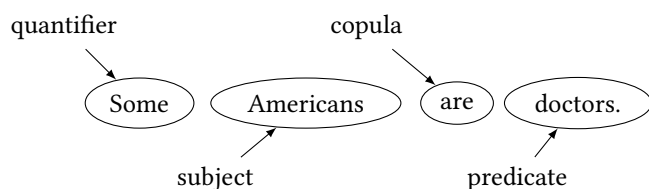


Figure 7.1: Parts of a quantified categorical statement.

In the previous chapter we noted that formal logic achieves content neutrality by replacing some or all of the ordinary words in a statement with symbols. For categorical logic, we are only going to be making one such substitution. Sometimes we will replace the classes referred to in a quantified categorical statement with capital letters that act as variables. Typically we will use the letter *S* when referring to the class in the subject term and *P* when referring to the predicate term, although sometimes more letters will be needed. Thus the sentence “Some Americans are doctors,”

above, will sometimes become “Some *S* are *P*.” The sentence “No dogs are cats” will sometimes become “No *S* is *P*.”

### *Quantity, Quality, Distribution, and Venn Diagrams*

Ordinary English contains all kinds of quantifiers, including the counting numbers themselves. In this chapter and the next, however, we are only going to deal with the quantifiers: “all,” “no,” and “some.” We are restricting ourselves to the quantifiers “all” and “some” because they are the ones that can easily be combined to create valid arguments using the system of logic that was invented by Aristotle. We will deal with other quantifiers in Chapter 13, on induction. There we will talk about much more specific quantified statements, like “Thirty percent of Canadians speak French,” and do a little bit of work with modern statistical methods.

The quantifier used in a statement is said to give the **QUANTITY** of the statement. Statements with the quantifier “All” are said to be “**UNIVERSAL**” and those with the quantifier “some” are said to be “**PARTICULAR**.”

Here “some” will just mean “at least one.” So, “some people in the room are standing” will be true even if there is only one person standing. Also, because “some” means “at least one,” it is consistent with the corresponding “all” statement. If I say “some people in the room are standing” it might actually be that *all* people in the room are standing, because if all people are standing, then at least one person is standing. This can sound a little weird, because in ordinary circumstances, you wouldn’t bother to point out that something applies to some members of a class when, in fact, it applies to all of them. It sounds odd to say “*some* dogs are mammals,” when in fact they *all* are. Nevertheless, when “some” means “at least one” it is perfectly true that some dogs are mammals.

In addition to talking about the quantity of statements, we will talk about their **QUALITY**. The quality of a statement refers to whether the statement is negated. Statements that include the words “no” or “not” are **NEGATIVE**, and other statements are **AFFIRMATIVE**. Combining quantity and quality gives us four basic types of quantified categorical statements, which we call the **STATEMENT MOODS** or just “moods.” The four moods are labeled with the letters A, E, I, and O. Statements that are universal and affirmative are **MOOD-A STATEMENTS**. Statements that are universal and negative are **MOOD-E STATEMENTS**. Particular and affirmative statements are **MOOD-I STATEMENTS**, and particular and negative statements are **MOOD-O STATEMENTS**. (See Table 7.1.)

Aristotle didn’t actually use those letters to name the kinds of categorical propositions. His later followers writing in Latin came up with the idea. They remembered the labels because the “A” and the “I” were in the Latin word “*affirmo*,” (“I affirm”) and the “E” and the “O” were in the Latin word “*nego*” (“I deny”).

<u>Mood</u>	<u>Form</u>	<u>Example</u>
A	All <i>S</i> are <i>P</i> .	All dogs are mammals.
E	No <i>S</i> are <i>P</i> .	No dogs are reptiles.
I	Some <i>S</i> are <i>P</i> .	Some birds can fly.
O	Some <i>S</i> are not <i>P</i> .	Some birds cannot fly.

Table 7.1: The four moods of categorical statements

The **DISTRIBUTION** of a categorical statement refers to how the statement describes its subject and predicate class. A term in a sentence is said to be distributed if a claim is being made about the whole class. In the sentence “All dogs are mammals,” the subject class, dogs, is distributed, because the quantifier “All” refers to the subject. The sentence is asserting that every dog out there is a mammal. On the other hand, the predicate class, mammals, is not distributed, because the sentence isn’t making a claim about all the mammals. We can infer that at least some of them are dogs, but we can’t infer that all of them are dogs. So in mood-A statements, only the subject is distributed.

On the other hand, in an *I* sentence like “Some birds can fly” the subject is not distributed. The quantifier “some” refers to the subject, and indicates that we are not saying something about all of that subject. We also aren’t saying anything about all flying things, either. So in mood-I statements, neither subject nor predicate is distributed.

Even though the quantifier always refers to the subject, the predicate class can be distributed as well. This happens when the statement is negative. The sentence “No dogs are reptiles” is making a claim about all dogs: they are all not reptiles. It is also making a claim about all reptiles: they are all not dogs. So mood-E statements distribute both subject and predicate. Finally, negative particular statements (mood-O) have only the predicate class distributed. The statement “some birds cannot fly” does not say anything about all birds. It does, however say something about all flying things: the class of all flying things excludes some birds.

The quantity, quality, and distribution of the four forms of a categorical statement are given in Table 7.2. The general rule to remember here is that universal statements distribute the subject, and negative statements distribute the predicate.

<b>Mood</b>	<b>Form</b>	<b>Quantity</b>	<b>Quality</b>	<b>Terms Distributed</b>
A	All <i>S</i> are <i>P</i>	Universal	Affirmative	S
E	No <i>S</i> are <i>P</i>	Universal	Negative	S and P
I	Some <i>S</i> are <i>P</i>	Particular	Affirmative	None
O	Some <i>S</i> are not <i>P</i>	Particular	Negative	P

Table 7.2: Quantity, quality, and distribution.

In 1880, the English logician John Venn published two essays on the use of diagrams with circles to represent categorical propositions.<sup>1</sup> Venn noted that the best use of such diagrams so far had come from the brilliant Swiss mathematician Leonhard Euler, but they still had many problems, which Venn felt could be solved by bringing in some ideas about logic from his fellow English logician George Boole. Although Venn only claimed to be building on the long logical tradition he traced, since his time these kinds of circle diagrams have been known as **VENN DIAGRAMS**.

<sup>1</sup> See Venn1880a, Venn1880b and

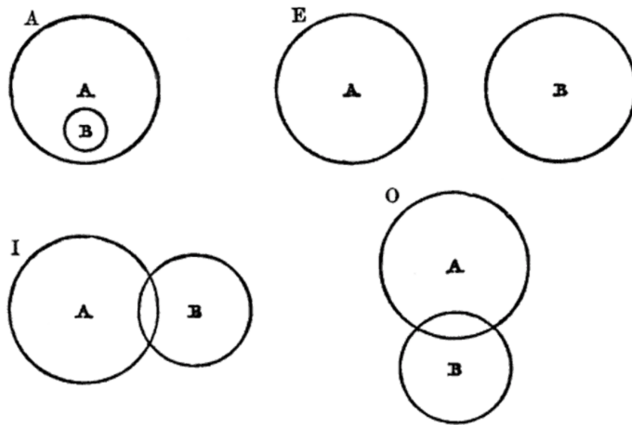


Figure 7.2: Euler diagrams from Scottish philosopher William Hamilton's *Lectures on Logic*.

In this section we are going to learn to use Venn diagrams to represent our four basic types of categorical statement. Later in this chapter, we will find them useful in evaluating arguments.

Let us start with a statement in mood A: “All *S* are *P*.” We are going to use one circle to represent *S* and another to represent *P*. There are a couple of different ways we could draw the circles if we wanted to represent “All *S* are *P*.” One option would be to draw the circle for *S* entirely inside the circle for *P*, as in the Euler diagram in the top-left of figure 7.2.

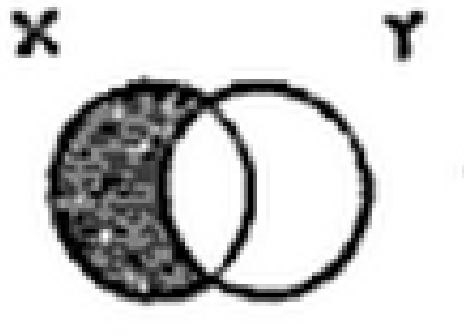


Figure 7.3: Venn's original diagram for an mood-A statement (Venn (?)). Screenshot from Google Books by J. Robert Loftis.

It is clear from the top left Euler diagram in figure 7.2 that all *B* are in

fact  $A$ . And outside of college logic classes, you may have seen people use a diagram like this to represent a situation where one group is a subclass of another. You may have even seen people call concentric circles like this a Venn diagram. But Venn did not think we should put one circle entirely inside the other if we just want to represent “All  $S$  is  $P$ .” Technically speaking figure 7.2 shows Euler circles, not Venn diagrams.

Venn pointed out that the circles in the top left of figure 7.2 don’t just say that “All  $B$  are  $A$ .” They also say that “All  $A$  are  $B$ ” is false. But we don’t necessarily know that if we have only asserted “All  $B$  are  $A$ .” The statement “All  $B$  are  $A$ ” leaves it open whether the  $B$  circle should be smaller than or the same size as the  $A$  circle.

Venn suggested that to represent just the content of a single proposition, we should always begin by drawing partially overlapping circles. This means that we always have spaces available to represent the four possible ways the terms can combine. Thus, for a single sentence we should always draw the diagram as in figure 7.4.

To return to our initial question, suppose we want to represent the sentence “All  $S$  is  $P$ ” using Venn diagrams. To do so we will shade the region of  $S$  that does not overlap with  $P$ . This tells us that the left-hand side of  $S$  is empty and so everything that has property  $S$  must also have property  $P$ . (See figure 7.5.)

If we want to say that something does exist in a region, we put an “ $x$ ” in it. The diagram for “Some  $S$  are  $P$ ” is in figure 7.6.

If a region of a Venn diagram is blank, if it is neither shaded nor has an  $x$  in it, it could go either way. Maybe such things exist, maybe they do not.

The Venn diagrams for all four basic forms of categorical statements are in figure ???. Notice that when we draw diagrams for the two universal forms,  $A$  and  $E$ , we do not draw any  $x$ ’s. For these forms we are only ruling out possibilities, not asserting that things actually exist. This is part of what Venn learned from Boole, and we will see its importance in Section ??.

Finally, notice that so far, we have only been talking about categorical statements involving the variables  $S$  and  $P$ . Sometimes, though, we will want to represent statements in regular English. To do this, we will include a key saying what the variables  $S$  and  $P$  represent in this case. We will call a list that assigns English phrases or sentences to variable names a **TRANSLATION KEY**. These are sometimes also called “symbolization keys” or simply just “dictionaries.” As our logical systems get more complicated, the symbolization keys will get more complicated. For now, though, they just consist of a note saying what the  $S$  and  $P$  stand for. For instance, this is the diagram for “No dogs are reptiles.”

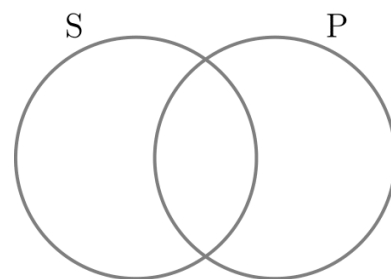


Figure 7.4: A basic Venn diagram.

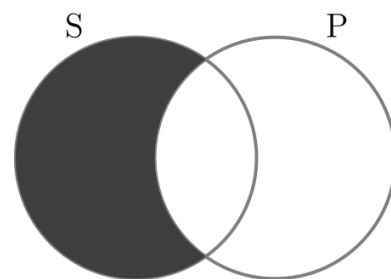


Figure 7.5: A Venn diagram for a mood-A sentence.

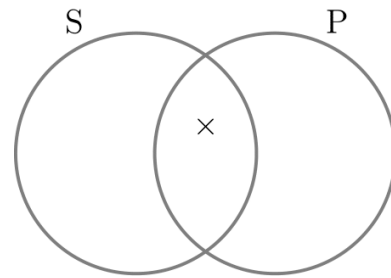


Figure 7.6: A Venn diagram for a mood-I sentence.

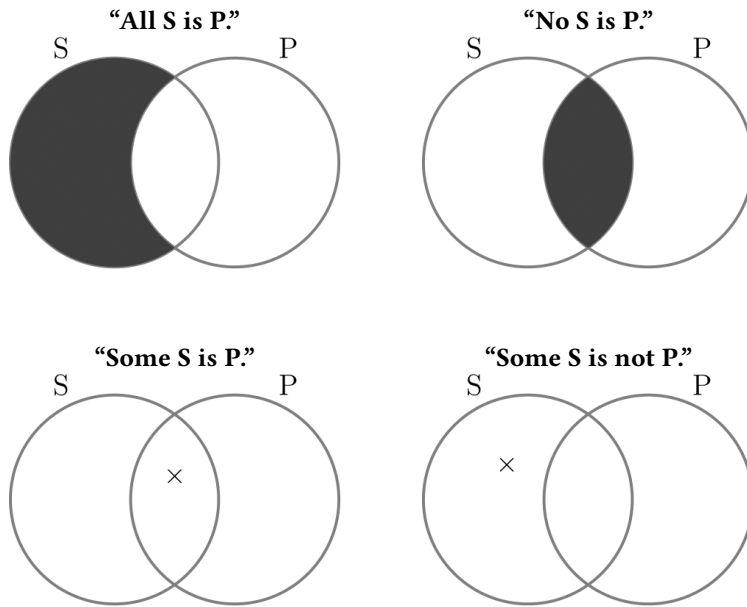


Figure 7.7: Venn diagrams for all four categorical moods.

### *Transforming English into Logically Structured English*

Because the four basic forms are stated using variables, they have a great deal of generality. We can expand on that generality by showing how many different kinds of English sentences can be represented as sentences in our four basic forms. We already touched on this a little in section 7, when we look at sentences like “Thirty percent of Canadians speak French.” There we saw that the predicate was not explicitly a class. We needed to change “speak French” to “people who speak French.” In this section, we are going to expand on that to show how ordinary English sentences can be transformed into something we will call “logically structured English.” **LOGICALLY STRUCTURED ENGLISH** is English that has been put into a standardized form that allows us to see its logical structure more clearly and removes ambiguity. Doing this is a step towards the creation of formal languages, which we will start doing in Chapter 10.

Transforming English sentences into logically structured English is fundamentally a matter of understanding the meaning of the English sentence and then finding the logically structured English statements with the same or similar meaning. Sometimes this will require judgment calls. English, like any natural language, is fraught with ambiguity. One of our goals with logically structured English is to reduce the amount of ambiguity. Clarifying ambiguous sentences will always require making judgments that can be questioned. Things will only get harder when we start using full blown formal languages in Chapter ??, which are supposed to be completely free of ambiguity.

To transform a quantified categorical statement into logically structured English, we have to put all of its elements in a fixed order and be sure they are all of the right type. All statements must begin with the quantifiers “All” or “Some” or the negated quantifier “No.” Next comes the subject term, which must be a plural noun, a noun phrase, or a variable that stands for any plural noun or noun phrase. Then comes the copula “are” or the negated copula “are not.” Last is the predicate term, which must also be a plural noun or noun phrase. We also specify that you can only say “are not” with the quantifier “some,” that way the universal negative statement is always phrased “No *S* are *P*,” not “All *S* are not *P*.” Taken together, these criteria define the **STANDARD FORM FOR A CATEGORICAL STATEMENT** in logically structured English.

The subsections below identify different kinds of changes you might need to make to put a statement into logically structured English. Sometimes translating a sentence will require using multiple changes.

### *Change the Predicate into a Noun Phrase*

In section 7 we saw that “Some Canadians speak French” has a verb phrase “speaks French” instead of a copula and a plural noun phrase. To transform these sentences into logically structured English, you need to add the copula and turn all the terms into plural nouns or plural noun phrases. Adding a plural noun phrase means you have to come up with some category, like “people” or “animals.” When in doubt, you can always use the most general category, “things.” Below are some examples

<u>English</u>	<u>Logically Structured English</u>
No cats bark.	No cats are animals that bark.
All birds can fly.	All birds are animals that can fly.
Some thoughts should be left unsaid.	Some thoughts are things that should be left unsaid.

Sometimes English sentences will have a copula and an adjective or adjective phrase as the predicate. These need to be changed to noun phrases, just as the verb phrases did.

<u>English</u>	<u>Logically Structured English</u>
Some roses are red.	Some roses are red flowers.
Football players are strong.	All football players are strong persons.
Some words are hurtful.	Some words are hurtful things.

Again, you will have to come up with a category for the predicate, and when it doubt, you can just use “things.”



*Standardize the Quantifier*

English has a wide variety of ways to express quantity. We need to reduce all of these to either “all” or “some,” plus negations. Here are some examples:

English

Most people with a PhD in psychology are female.  
 Among the things that Sylvia inherited was a large mirror  
 There are Americans that are doctors.  
 At least a few Americans are doctors.  
 A man is walking down the street.  
 Every day is a blessing.  
 Whatever is a dog is not a cat.  
 Not a single dog is a cat.  
 Take nothing for granted  
 Something is rotten in Denmark  
 Everything is coming up roses  
 “What does not destroy me, makes me stronger.” –  
 Friedrich Nietzsche

Logically Structured English

Some people with a PhD in psychology are female.  
 Some things that Sylvia inherited were large mirrors  
 Some Americans are doctors.  
 Some Americans are doctors.  
 Some men are things that are walking down the street.  
 All days are blessings.  
 No dogs are cats.  
 No dogs are cats.  
 No things are things that should be taken for granted  
 Some things are things that are rotten in Denmark  
 All things are things that are coming up roses  
 All things that do not destroy me are things that make me stronger.

Notice in the last case we are losing quite a bit of information when we transform the sentence into logically structured English. “Most” means more than fifty percent, while “some” could be any percentage less than a hundred. This is simply a price we have to pay in creating a standard logical form. As we will see when we move to constructing artificial languages in Chapter ??, no logical language has the expressive richness of a natural language.

Sometimes universal statements in English don’t have an explicit quantifier. Instead they use a plural noun or indefinite article to express generality.

English

Boots are footwear.  
 Giraffes are tall.  
 A dog is not a cat.  
 A lion is a fierce creature.

Logically Structured English

All boots are footwear.  
 All giraffes are tall things.  
 No dogs are cats.  
 All lions are fierce creatures.

Notice that in the second sentence we had to make two changes, adding both the words “All” and “things.”

In the last two sentences, the indefinite article “a” is being used to create a kind of generic sentence. Not all sentences using the indefinite article work this way. The list before this one included the example “A man

is walking down the street.” This sentence is not talking about all men generically. It is talking about a specific man whose identity is unknown. Here the indefinite article is being used like a nonstandard version of the quantifier “some,” which is why it appeared in the earlier list. You will have to use your good judgment and understanding of context to know when the indefinite article is being used like the word “all” and when it is being used like the word “some.”

You may be someone who interprets these indefinite sentences slightly differently. Sometimes, English speakers say things like, “Dogs bark.” to mean something like, “*Most* dogs are things that bark.” or “Typically, dogs are things that bark.” Indefinite articles and generic sentences like this are quintessentially *vague* sentences. Because we do not have a way to formalize the quantifier ‘most’ we are going to translate generics as expressing a universal quantity. So, “Dogs bark.” will become “All dogs are things that bark.” More advanced categorical logics that have an infinite number of quantifiers attempt to translate the detail of these generics more precisely, but we won’t worry about that for now.

English also uses specialized adverbial phrases as quantifiers for people, places and times. If we want to talk about all people, we use a specialized quantifier like “everyone,” “someone” or “no one.” We use “everywhere,” “somewhere,” and “nowhere” for places, and “always,” “sometimes,” and “never” for times. All of these need to be transformed into using the simple quantifiers “all” or “some,” plus negations.

<u>English</u>	<u>Logically Structured English</u>
Someone in America is a doctor.	Some Americans are doctors.
Not everyone who is an adult is a logician.	Some adults are not logicians.
“Whenever you need me, I’ll be there.” – Michael Jackson	All times that you need me are times that I will be there.
“We are never, ever, ever getting back together.” – Taylor Swift	No times are times when we will get back together.
“Whoever fights with monsters should be careful lest he thereby become a monster.” –Friedrich Nietzsche	All persons who fight with monsters are persons who should be careful lest they become a monster.

### *Standardize Alternative Universal Forms*

Many constructions in English can be represented as universal statements in Logically Structured English, either affirmative (A) or negative (E)

For instance, it turns out that statements about individual people or specific objects can be represented by A or E statements. This is not something Aristotle originally noticed. For him a statement like “Socrates is mortal,” for Aristotle, were neither universal nor particular. They were a third class

he called “singular.” The power of categorical logic was expanded considerably when it was realized singular statements can be converted into universal statements. The trick is to add a phrase like “All things identical to...” to our singular sentence. Essentially we are adding a universal quantifier that only picks out one specific object.

#### English

Socrates is mortal.

The Empire State Building is tall.

Ludwig was not happy.

#### Logically Structured English

All persons identical with Socrates are mortal.

All things identical to The Empire State Building are tall things.

No people identical with Ludwig are happy people.

Another kind of statement that can be transformed into a universal statement is a conditional. A conditional is a statement of the form “If ... then ...” They will become a big focus of our attention starting in Chapter ?? when we begin introducing modern formal languages. They are not given special treatment in the Aristotelian tradition, however. Instead, where we can, we just treat them as categorical generalizations:

#### English

If something is a cat, then it is a feline.

If something is a dog, then it's not a cat.

#### Logically Structured English

All cats are feline.

No dogs are cats.

The word “only” is used in a couple of different constructions in English that can be represented as universal statements. The first kind are called “exclusive propositions.” These are statements that say the subject excludes everything except what is in the predicate. For instance the sentence “Only people over 21 may drink” says that the class of people who may drink excludes everyone except those who are over 21. In English exclusive propositions are created using the words “only,” “none but,” or “none except.” These statements become A statements when translated into logically structured English. So “Only people over 21 may drink” becomes “If you may drink, you are over 21.” It is important to see that in each case these words are used to introduce the predicate, not the subject. In the sentence “Only people over 21 may drink,” the term “people over 21” is actually the predicate, and “people who may drink” is the subject.

#### English

Only people over 21 may drink.

No one, except those with a ticket, may enter the theater.

None but the strong survive.

#### Logically Structured English

All people who drink are over 21.

All people who enter the theater have a ticket.

All people who survive are strong people.

Sentences with “The only” are a little different than sentences regular exclusive propositions, which just have “only” in them. The sentence “Humans are the only animals that talk on cell phones” should be translated as “All animals who talk on cell phones are humans.” In this sentence, “the only” introduces the subject, rather than the predicate. The statement still asserts that the subject excludes everything except what is in the predicate, and we still represent them using mood A statements.

#### English

Humans are the only animals who talk on cell phones.

Shrews are the only venomous mammal in North America.

#### Logically Structured English

All animals who talk on cell phones are human.

All venomous mammals in North America are shrews.

Transforming sentences into [logically structured English](#) requires judgment and attention to the nuances of meaning in English. You must be able to recognize which of the transformations describe above needs to be applied and apply it correctly. One frequent mistake by people starting out is to overgeneralize. We saw at the start of the subsection on alternative universal forms that singular propositions can be turned into universal propositions by adding the phrase “Things identical to ...” Once you get in the habit of doing this, it becomes tempting to add the phrase “things identical to ...” to everything, even when it isn’t necessary or doesn’t make sense. The sentence “Fido is a dog” should become “all things identical to Fido are dogs” in logically structured English, because “Fido” is a singular term referring to an individual dog. But with the sentence “dogs are mammals,” you do not need to add the phrase “All things identical to...”, because “dogs” is already a collective noun, not an individual.

The same is true for the phrases we use to transform adjective and verb phrases into noun phrases. The sentence “No cats bark” has to be changed, because “bark” is a verb, so it becomes “No cats are animals that bark” in Logically Structured English. But the sentence “No cats are reptiles” already has a noun, “reptiles,” for a predicate, so you do not need to transform it into “No cats are animals that are reptiles.” The key is not only knowing when to use the transformations we describe, but knowing when not to use them.

### *Key Terms*

- |                           |                          |
|---------------------------|--------------------------|
| 1. <b>Affirmative</b>     | 4. <b>Contraposition</b> |
| 2. <b>Complement</b>      | 5. <b>Contraries</b>     |
| 3. <b>Contradictories</b> | 6. <b>Converse</b>       |

- |                                      |   |
|--------------------------------------|---|
| 7. Copula                            | 21. Quantifier                          |
| 8. Distribution                      | 22. Quantity                            |
| 9. Existential import                | 23. Square of opposition                |
| 10. Logically structured English     | 24. Standard form categorical statement |
| 11. Mood-A statement                 | 25. Subalternation                      |
| 12. Mood-E statement                 | 26. Subcontraries                       |
| 13. Mood-I statement                 | 27. Subject class                       |
| 14. Mood-O statement                 | 28. Translation key                     |
| 15. Negative                         | 29. Truth value                         |
| 16. Obverse                          | 30. Universal                           |
| 17. Particular                       | 31. Vacuous truth                       |
| 18. Predicate class                  | 32. Venn diagram                        |
| 19. Quality                          |   |
| 20. Quantified categorical statement |   |



## 8

# Categorical Syllogisms

### *Standard Form, Mood, and Figure*

So far we have just been looking at very short arguments using categorical statements. The arguments just had one premise and a conclusion that was often logically equivalent to the premise. For most of the history of logic in the West, however, the focus has been on arguments that are a step more complicated called **CATEGORICAL SYLLOGISMS**. A categorical syllogism is a two-premise argument composed of categorical statements. Aristotle began the study of this kind of argument in his book the *Prior Analytics*.<sup>1</sup> This work was refined over the centuries by many thinkers in the Pagan, Christian, Jewish, and Islamic traditions until it reached the form it is in today.

<sup>1</sup> A digital copy of the *Prior Analytics* is available at <http://classics.mit.edu/Aristotle/prior.html>

There are actually many kinds of two-premise arguments using categorical statements, but Aristotle only looked at arguments where each statement is in one of the moods A, E, I, or O. The arguments also had to have exactly three terms, arranged so that any two pairs of statements will share one term. Let's call a categorical syllogism that fits this more narrow description an **ARISTOTELIAN SYLLOGISM**. Here is a typical Aristotelian syllogism using only mood-A sentences:

(A1) All mammals are vertebrates.

(A2) All dogs are mammals.

(A3) ∴ All dogs are vertebrates.

Notice how the statements in this argument overlap each other. Each statement shares a term with the other two. Premise 2 shares its subject term with the conclusion and its predicate with Premise 1. Thus there are only three terms spread across the three statements. Aristotle dubbed these the major, middle, and minor premises, but there was initially some confusion about how to define them. In the 6th century, the Christian philosopher John Philoponus, drawing on the work of his pagan teacher Ammonius, decided to arbitrarily designate the **MAJOR TERM** as the predicate of the conclusion, the **MINOR TERM** as the subject of the conclusion, and

the **MIDDLE TERM** as the one term of the Aristotelian syllogism that does not appear in the conclusion. So in the argument above, the major term is “vertebrate,” the middle term is “mammal,” and the minor term is “dog.” We will adopt this convention as well. We can also define the **MAJOR PREMISE** as the one premise in an Aristotelian syllogism that names the major term, and the **MINOR PREMISE** as the one premise that names the minor term. So in the argument above, Premise 1 is the major premise and Premise 2 is the minor premise.

With these definitions in place, we can now define the **STANDARD FORM** in logically structured English. Recall that in section 7, we started standardizing our language into something we called “logically structured English” in order to remove ambiguity and to make its logical structure clear. The first step was to define the standard form for a categorical statement, which we did on page ?? . Now we do the same thing for an Aristotelian syllogism. We say that an Aristotelian syllogism is in standard form for logically structured English if and only if these criteria have been met: (1) all of the individual statements are in standard form, (2) each instance of a term is in the same format and is used in the same sense, (3) the major premise appears first, followed by the minor premise, and then the conclusion.

Once we standardize things this way, we can actually catalog every possible form of an Aristotelian syllogism. To begin with, each of the three statements can take one of four forms: A, E, I, or O. This gives us  $4 \times 4 \times 4$ , or 64 possibilities. These 64 possibilities are called the **SYLLOGISM MOOD**, and we designate it just by writing the three letters of the moods of the statements that make it up. So the mood of the argument on page 111 is simply AAA. All 64 moods are listed in table 8.1, for reference.

AAA	AAE	AAI	AAO
AEA	AEE	AEI	AEO
AIA	AIE	AII	AIO
AOA	AOE	AOI	AOO
EAA	EAE	EAI	EAO
EEA	EEE	EEI	EEO
EIA	EIE	EII	EIO
EOA	EOE	EOI	EOO
IAA	IAE	IAI	IAO
IEA	IEE	IEI	IEO
IIA	IIE	III	IIO
IOA	IOE	IOI	IOO
OAA	OAE	OAI	OAO
OEA	OEE	OEI	OEO
OIA	OIE	OII	OIO
OOA	OOE	OOI	OOO

Table 8.1: The sixty-four moods of the Aristotelian syllogism

In addition to varying the kind of statements we use in an Aristotelian



sylllogism, we can also vary the placement of the major, middle, and minor terms. There are four ways we can arrange them that fit the definition of an Aristotelian syllogism in standard form, shown in Table 8.2. Here  $P$  stands for the major term,  $S$  for the minor term, and  $M$  for the middle. The thing to pay attention to is the placement of the middle terms. In figure 1, the middle terms form a line slanting down to the right. In figure 2, the middle terms are both pushed over to the right. In figure 3, they are pushed to the left, and in figure 4, they slant in the opposite direction from figure 1.

(B1) **M**  $P$   
 (B2)  $S$  **M**  
 (B3)  $\therefore S$   $P$

Figure 1

(C1)  $P$  **M**  
 (C2)  $S$  **M**  
 (C3)  $\therefore S$   $P$

Figure 2

(D1) **M**  $P$   
 (D2) **M**  $S$   
 (D3)  $\therefore S$   $P$

Figure 3

(E1)  $P$  **M**  
 (E2) **M**  $S$   
 (E3)  $\therefore S$   $P$

Figure 4

Table 8.2: The four figures of the Aristotelian syllogism

The combination of 64 moods and 4 figures gives us a total of 256 possible Aristotelian syllogisms. We can name them by simply giving their mood and figure. So this is OAO-3:

(F1) Some  $M$  are not  $P$ .  
 (F2) All  $M$  are  $S$ .  
 (F3)  $\therefore$  Some  $S$  are not  $P$ .

Syllogism OAO-3 is a valid argument. We will be able to prove this with Venn diagrams in the next section. For now just read it over and try to see intuitively why it is valid. Most of the 256 possible syllogisms, however, are not valid. In fact, most of them, like IIE-2, are quite obviously invalid:

(G1) Some  $P$  are  $M$ .  
 (G2) Some  $S$  are  $M$ .  
 (G3)  $\therefore$  No  $S$  are  $P$ .

Given an Aristotelian syllogism in ordinary English, we can transform it into standard form in logically structured English and identify its mood and figure. Consider the following:

No geckos are cats. I know this because all geckos are lizards, but cats aren't lizards.

The first step is to identify the conclusion, using the basic skills you acquired back in Chapter 1. In this case, you can see that “because” is a premise indicator word, so the statement before it, “No geckos are cats,” must be the conclusion.

Step two is to identify the major, middle, and minor terms. Remember that the major term is the predicate of the conclusion, and the minor term is the subject. So here the major term is “cats,” the minor term is “geckos.” The leftover term, “lizards,” must then be the middle term.

We show that we have identified the major, middle, and minor terms by writing a [translation key](#). As we saw on page 159 a translation key is just a list that assigns English phrases or sentences to variable names. For categorical syllogism, this means matching the English phrases for the terms with the variables  $S$ ,  $M$ , and  $P$ .

$S$  (*minor term*): Geckos

$M$  (*middle term*): Lizards

$P$  (*major term*): Cats

Step three is to write the argument in canonical form using variables for the terms. The last statement, “cats aren’t lizards,” is the major premise, because it has the major term in it. We need to change it to standard form, however, before we substitute in the variables. So first we change it to “No cats are lizards.” Then we write “No  $S$  are  $M$ .” For the minor premise and the conclusion we can just substitute in the variables, so we get this:

(H1) No  $P$  are  $M$ .

(H2) All  $S$  are  $M$ .

(H3)  $\therefore$  No  $S$  are  $P$ .

Step four is to identify mood and figure. We can see that this is figure 2, because the middle term is in the predicate of both premises. Looking at the form of the sentences tells us that this is EAE.

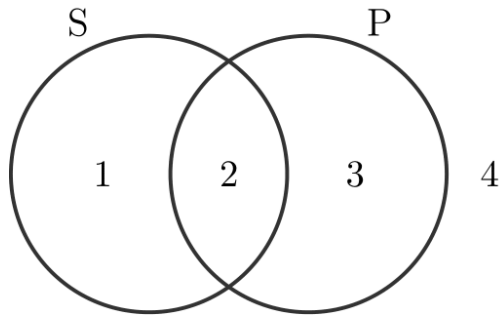
### *Testing Validity*

We have seen that there are 256 possible categorical arguments that fit Aristotle’s requirements. Most of them are not valid. In this section, we will learn to use Venn diagrams to sort the valid arguments from the invalid. The method we will use is simply an extension of what we did in the last chapter, except with three circles instead of two.

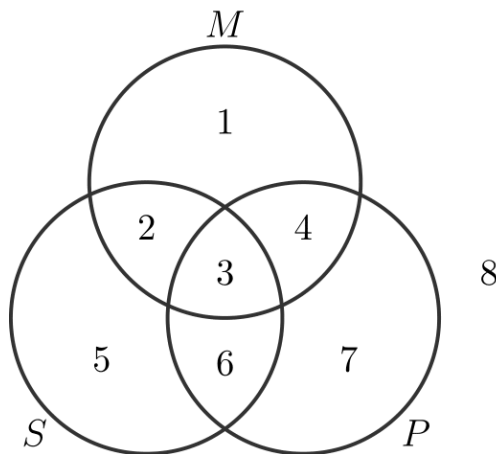
### *Venn Diagrams for Single Propositions*

In the previous chapter, we drew Venn diagrams with two circles for arguments that had had two terms. The circles partially overlapped, giving us

four areas, each of which represented a way an individual could relate to the two classes. So area 1 represented things that were *S* but not *P*, etc.



Now that we are considering arguments with three terms, we will need to draw three circles, and they need to overlap in a way that will let us represent the eight possible ways an individual can be inside or outside these three classes.



So in this diagram, area 1 represents the things that are *M* but not *S* or *P*, area 2 represents the things that are *M* and *S* but not *P*, etc. We'll mostly ignore region 8 as things that are neither *S*, nor *M*, nor *P*.

As before, we represent universal statements by filling in the area that the statement says cannot be occupied. The only difference is that now there are more possibilities. So, for instance, there are now four mood-A propositions that can occur in the two premises. The major premise can either be “All *P* are *M*” or “All *M* are *P*,” and the minor premise can be either “All *S* are *M*” or “All *M* are *S*.” The Venn diagrams for those four sentences are given in figure 8.1.

Similarly, there are four mood-E propositions that can occur in the premises of an Aristotelian syllogism: “No *P* are *M*,” “No *M* are *P*,” “No *S* are *M*,” and “No *M* are *S*.” And again, we diagram these by shading out overlap between the two relevant circles. In this case, however, the first two

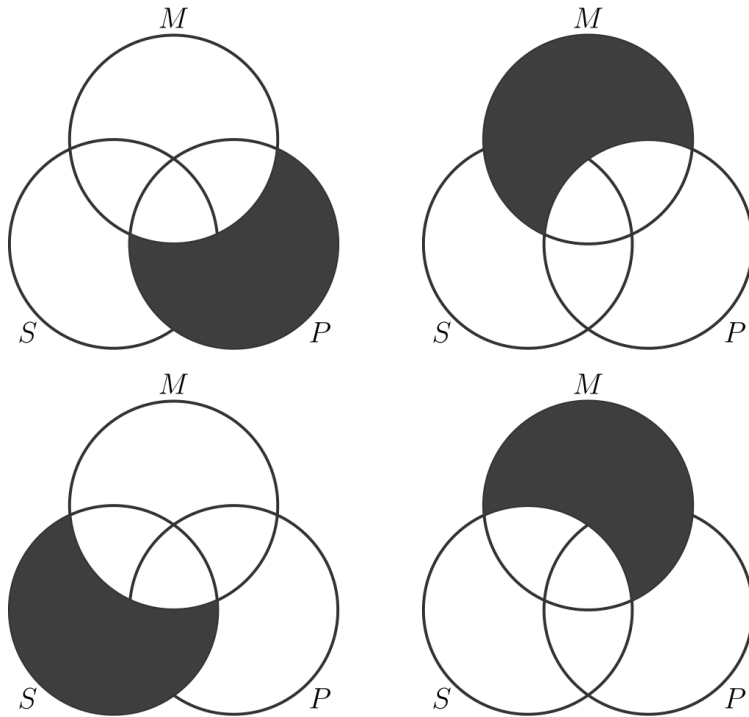


Figure 8.1: The four Venn diagrams for mood A sentences.

statements are equivalent by conversion, as are the second two. Thus we only have two diagrams to worry about. (See figure 8.2.)

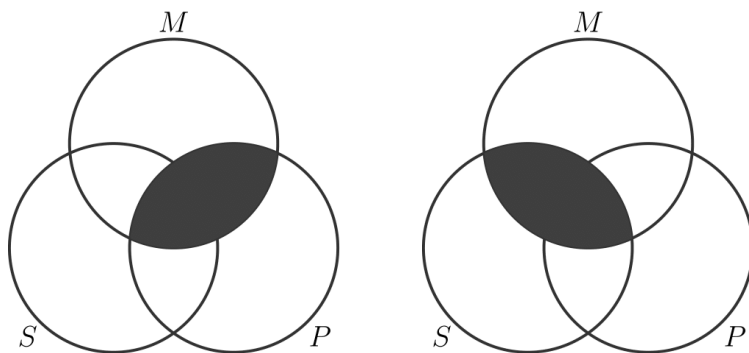
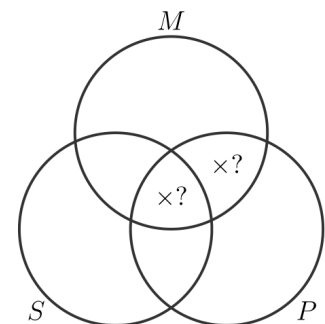
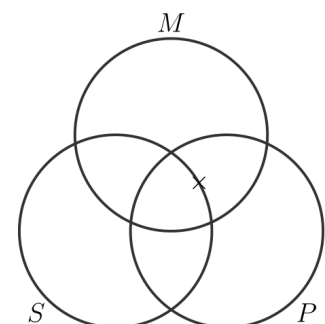


Figure 8.2: The two Venn diagrams for mood E sentences.

Particular propositions are a bit trickier. Consider the statement “Some  $M$  are  $P$ .” With a two circle diagram, you would just put an  $x$  in the overlap between the  $M$  circle and the  $P$  circle. But with the three circle diagram, there are now two places we can put it. It can go in either area 3 or area 4, as shown in figure 8.3.

The solution here will be to put the  $x$  on the boundary between areas 3 and 4, to represent the fact that it *could* go in either location. An example of this is shown in figure 8.4.

Sometimes, however, you won’t have to draw the  $X$  on a border between

Figure 8.3: It is ambiguous where the  $X$  should go on this Venn diagram.Figure 8.4: We place the  $X$  on the border between these regions to indicate that it *could* be in either.

two areas, because you will already know that one of those areas can't be occupied. Suppose, for instance, that you want to diagram "Some  $M$  are  $P$ ," but you already know that all  $M$  are  $S$ . You would diagram "All  $M$  are  $S$ " as in figure 8.5

Then, when it comes time to add the  $x$  for "Some  $M$  are  $P$ ," you know that it has to go in the exact center of the diagram. This is shown in figure 8.6

The Venn diagrams for the particular premises that can appear in Aristotelian syllogisms are given in figure 8.7. The figure assumes that you are just representing the individual premises, and don't know any other premises that would shade some regions out. Again, some of these premises are equivalent by conversion, and thus share a Venn diagram.

### *Venn Diagrams for Full Syllogisms*

In the last chapter, we used Venn diagrams to evaluate arguments with single premises. It turned out that when those arguments were valid, the conclusion was logically equivalent to the premise, so they had the exact same Venn diagram. This time we have two premises to diagram, and the conclusion won't be logically equivalent to either of them. Nevertheless we will find that for valid arguments, once we have diagrammed the two premises, we will also have diagrammed the conclusion.

First we need to specify a rule about the order to diagram the premises in: if one of the premises is universal and the other is particular, diagram the universal one first. This will allow us to narrow down the area where we need to put the  $x$  from the particular premise, as in the example above where we diagrammed "Some  $M$  are  $P$ " assuming that we already knew that all  $M$  are  $S$ .

Let's start with a simple example, an argument with the form AAA-1.

- (I1) All  $M$  are  $P$ .
- (I2) All  $S$  are  $M$ .
- (I3)  $\therefore$  All  $S$  are  $P$ .

Since both premises are universal, it doesn't matter what order we do them in. Let's do the major premise first. The major premise has us shade out the parts of the  $M$  circle that don't overlap the  $P$  circle, as shown in figure 8.8

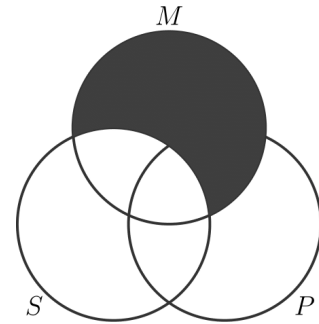


Figure 8.5: The  $x$  is forced into region 3 because we know that nothing can be in region 4.

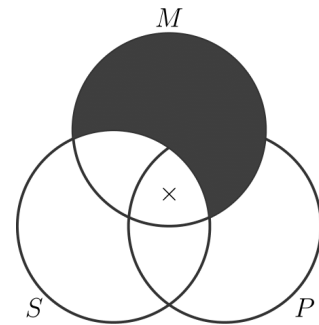


Figure 8.6: The  $x$  is forced into region 3 because we know that nothing can be in region 4.

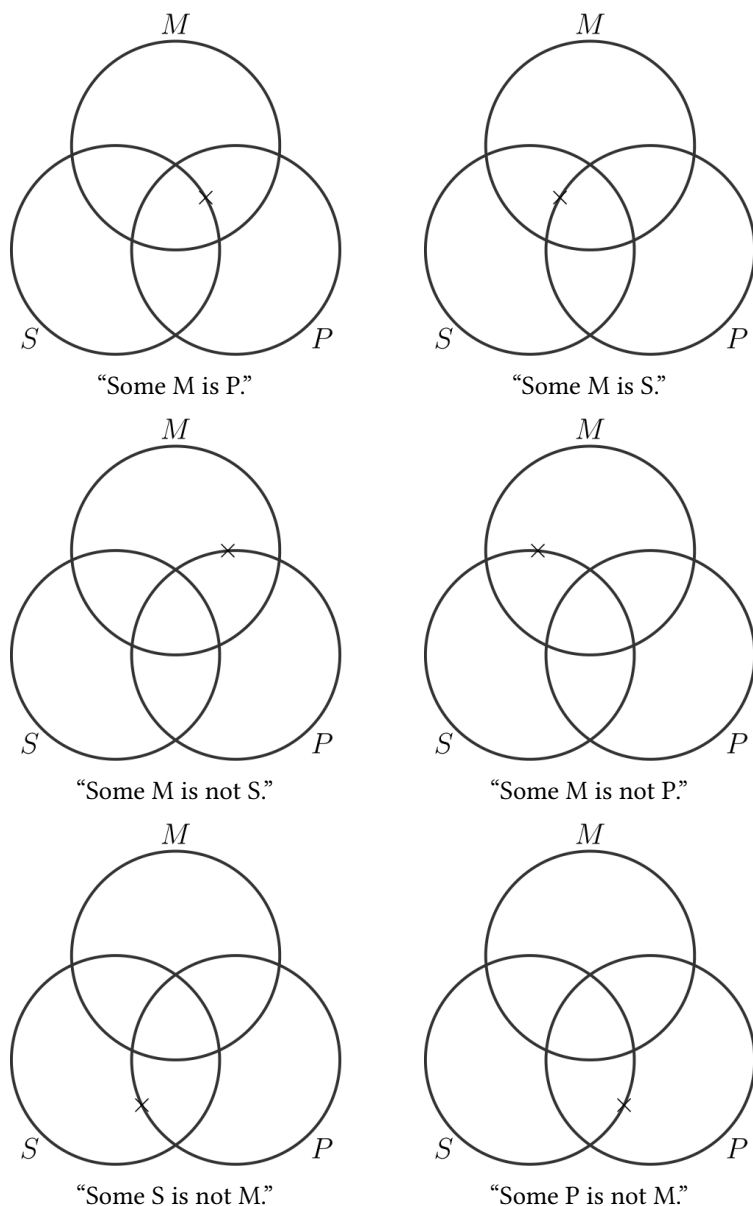


Figure 8.7: Basic Venn diagrams for all particular categorical sentences. Note that syllogisms with universal sentences may shade some regions and force the X off the border and completely into one of the regions.

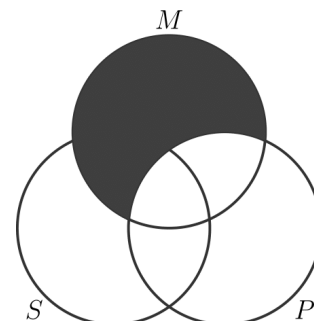


Figure 8.8: "All  $M$  are  $P$ ."

The second premise, on the other hand, tells us that there is nothing in the  $S$  circle that isn't also in the  $M$  circle. We put that together with the first diagram, and we get the Venn diagram in figure 8.9:

From this we can see that the conclusion must be true. All  $S$  are  $P$ , because the only space left in  $S$  is the area in the exact center, area 7.

Now let's look at an argument that is invalid. One of the interesting things about the syllogism AAA-1 is that if you change the figure, it ceases to be valid. Consider AAA-2.

(J1) All  $P$  are  $M$ .

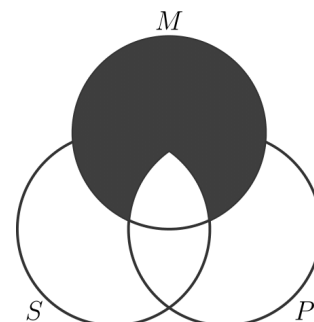


Figure 8.9: "All  $M$  are  $P$ ." and "All  $S$  are  $M$ ."

- (J2) All  $S$  are  $M$ .  
 (J3)  $\therefore$  All  $S$  are  $P$ .

Again, both premises are universal, so we can do them in any order, so we will do the major premise first. This time, the major premise tells us to shade out the part of  $P$  that does not overlap  $M$  as shown in figure 8.10

The second premise adds the idea that all  $S$  are  $M$ , which we diagram as shown in figure 8.11.

Now we ask if the diagram of the two premises also shows that the conclusion is true. Here the conclusion is that all  $S$  are  $P$ . If this diagram had made this true, we would have shaded out all the parts of  $S$  that do not overlap  $P$ . But we haven't done that. It is still possible for something to be in area 2. Therefore, this argument is invalid.

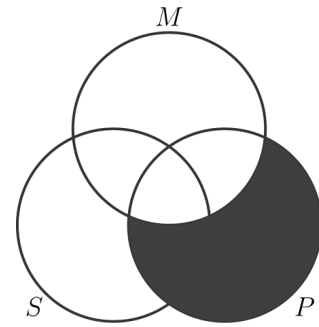


Figure 8.10: "All  $P$  are  $M$ ."

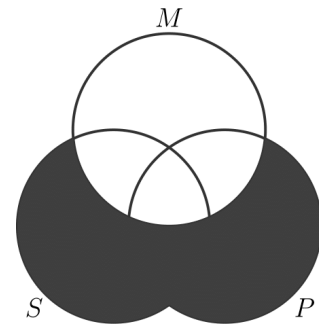


Figure 8.11: "All  $P$  are  $M$ ." and "All  $S$  are  $M$ ."

Now let's try an argument with a particular statement in the premises. Consider the argument IAI-1:

- (K1) Some  $M$  are  $P$ .  
 (K2) All  $S$  are  $M$ .  
 (K3)  $\therefore$  Some  $S$  are  $P$ .

Here, the second premise is universal, while the first is particular, so we begin by diagramming the universal premise as shown in figure 8.12

Then we diagram the particular premise "Some  $M$  are  $P$ ." This tells us that something is in the overlap between  $M$  and  $P$ , but it doesn't tell us whether that thing is in the exact center of the diagram or in the area for things that are  $M$  and  $P$  but not  $S$ . Therefore, we place the  $X$  on the border between these two areas as shown in figure 8.13

Now we can see that the argument is not valid. The conclusion asserts that something is in the overlap between  $S$  and  $P$ . But the  $X$  we drew does not necessarily represent an object that exists in that overlap. There is something out there that could be in area 3, but it could just as easily be in area 4. The second premise doesn't help us, because it just rules out the existence of objects in areas 5 and 6.

For a final example, let's look at a case of a valid argument with a particular statement in the premises. If we simply change the figure of the argument in the last example from 1 to 3, we get a valid argument. This is the argument IAI-3:

- (L1) Some  $M$  are  $P$ .  
 (L2) All  $M$  are  $S$ .  
 (L3)  $\therefore$  Some  $S$  are  $P$ .

Again, we begin with the universal premise. This time it tells us to shade out part of the  $M$  circle as shown in figure 8.14

But now we fill in the parts of  $M$  that don't overlap with  $S$ , we have to put the  $X$  in the exact center of the diagram.

And now this time we see that "Some  $S$  are  $P$ " has to be true based on the premises, because the  $X$  has to be in area 3. So this argument is valid.

Using this method, we can show that only 15 of the 256 possible syllogisms are valid. These, it turns out, are the only unconditionally valid syllogisms! Remember, however, that the Venn diagram method uses *Boolean assumptions about existential import*. This means that the Venn diagram method does not assume that an object *actually exists* in a region unless the presence of an  $X$  makes such existence explicit.

If you make additional assumptions about existential import you will allow more valid syllogisms. The next chapter on categorical logic will discuss the existential import assumption and the "Aristotelian interpretation" of syllogisms. The additional syllogisms we will be able to prove valid in

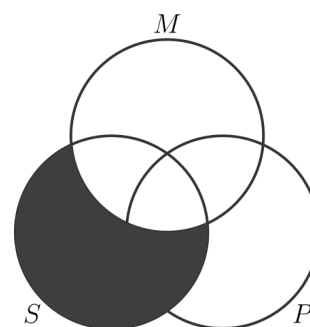


Figure 8.12: "All  $S$  are  $M$ ."

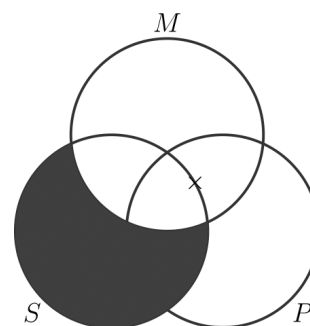


Figure 8.13: "Some  $M$  are  $P$ ." and "All  $S$  are  $M$ ."

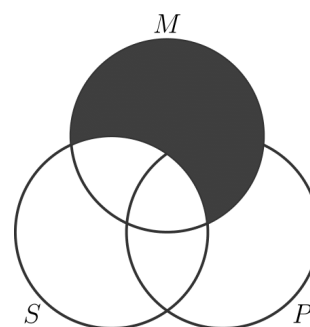


Figure 8.14: "All  $M$  are  $S$ ."

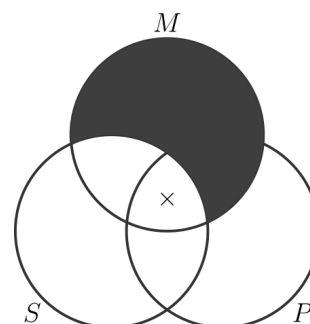


Figure 8.15: "Some  $M$  are  $P$ ." and "All  $M$  are  $S$ ."



the next chapter will be said to have **CONDITIONAL VALIDITY** because they are valid on the condition that the objects talked about in the universal statements actually exist. The 15 syllogisms that we can prove valid using the Venn diagram method have **UNCONDITIONAL VALIDITY**.

These syllogisms are given in Table 8.3.

Figure 1	Figure 2	Figure 3	Figure 4
Barbara (AAA)	Camestres (AEE)	Disamis (IAI)	Calemes (AEE)
Celarent (EAE)	Cesare (EAE)	Bocardo (OAO)	Dimatis (IAI)
Ferio (EIO)	Festino (EIO)	Ferison (EIO)	Fresison (EIO)
Darii (AII)	Baroco (AOO)	Datisi (AII)	

Table 8.3: The 15 unconditionally valid syllogisms.

The names on Table 8.3 come from the Christian part of the Aristotelian tradition, where thinkers were writing in Latin. Students in that part of the tradition learned the valid forms by giving each one a name. The vowels in the name represented the mood of the syllogism. So *Barbara* has the mood AAA, *Fresison* has the mood EIO, etc. The consonants in each name were also significant: they related to a process the Aristotelians were interested in called reduction, where arguments in the later figures were shown to be equivalent to arguments in the first figure, which was taken to be more self-evident. We won't worry about reduction in this textbook, however. The names of the valid syllogisms were often worked into a mnemonic poem.

The oldest known version of the poem appears in a late 13th century book called *Introduction to Logic* by William of Sherwood. Figure 8.16 is an image of the oldest surviving manuscript of the poem, digitized by the Bibliothèque Nationale de France.

The columns in Table 8.3 represent the four figures. Syllogisms with the same mood also appear in the same row. So the EIO sisters—*Ferio*, *Festino*, *Ferison*, and *Fresison*—fill up row 3. *Camestres* and *Calemes* share row 1; *Celarent* and *Cesare* share row 2; and *Darii* and *Datisi* share row 4.

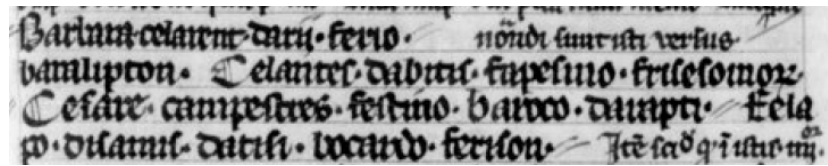


Figure 8.16: The oldest surviving version of the “Barbara, Celarent...” poem, from William of Sherwood. The manuscript is held at the Bibliothèque Nationale de France, ms. Lat. 16617, <http://gallica.bnf.fr/ark:/12148/btv1b9066740r>

### Key Terms

1. Aristotelian syllogism
2. Categorical syllogism
3. Conditional validity

4. **Critical term**
5. **Fallacy of exclusive premises**
6. **Fallacy of illicit process**
7. **Fallacy of particular premises**
8. **Fallacy of the undistributed middle**
9. **Major premise**
10. **Major term**
11. **Middle term**
12. **Minor premise**
13. **Minor term**
14. **Negative-affirmative fallacy**
15. **Standard form for an Aristotelian syllogism**
16. **Syllogism mood**
17. **Translation key**
18. **Unconditional validity**

## *Existential Import*

### *Existential Import and Conditionally Valid Forms*

In the last section, we mentioned that you can prove more syllogisms valid if you make different assumptions about existential import. Recall that a statement has existential import if, when you assert the statement, you are also asserting the existence of the things the statement talks about. (See page ??.) So if you interpret a mood-A statement as having existential import, it not only asserts “All *S* is *P*,” it also asserts “There are some *S*.” or, equivalently, “Some *S* exist.” Thus the mood-A statement “All unicorns have one horn” is false, if it is taken to have existential import, because unicorns do not exist. It is probably true, however, if you do not imagine the statement as having existential import. If anything is true of unicorns, it is that they would have one horn if they existed.<sup>1</sup>

We saw in section ?? that before Boole, Aristotelian thinkers had all sorts of opinions about existential import, or, as they put it, whether a term “supposits.” This generally led them to recognize additional syllogism forms as valid. You can see this pretty quickly if you just remember the traditional square of opposition. The traditional square allowed for many more valid immediate inferences than the modern square. It stands to reason that traditional ideas about existential import will also allow for more valid syllogisms.

Our system of Venn diagrams can’t represent all of the alternative ideas about existential import. For instance, it has no way of representing Ockham’s belief that mood-O statements do *not* have existential import. Nevertheless, it would be nice if we could expand our system of Venn diagrams to show that some syllogisms are valid if you make additional assumptions about existence.

Consider the argument Barbari (AAI-1).

(A1) All *M* are *P*.

(A2) All *S* are *M*.

(A3) ∴ Some *S* are *P*.

<sup>1</sup> Existence is a complex topic in philosophy, including the truth conditions for evaluating sentences like “Frodo Baggins is a hobbit.” or “Unicorns have one horn.” See e.g. the Stanford Encyclopedia of Philosophy article on Fictional Entities: <https://plato.stanford.edu/entries/fictional-entities/>

You won't find this argument in the list of unconditionally valid forms in Table 8.3. This is because under Boolean assumptions about existence it is not valid. The Venn diagram, which follows Boolean assumptions, shows this.

This is essentially the same argument as Barbara, but the mood-A statement in the conclusion has been replaced by a mood-I statement. We can see from the diagram that the mood-A statement "All *S* are *P*" is true. There is no place to put an *S* other than in the overlap with *P*. But we don't actually know the mood-I statement "Some *S* is *P*," because we haven't drawn an *x* in that spot. Really, all we have shown is that *if* an *S* existed, it would be *P*.

But by the traditional square of opposition (p. ??) we know that the mood-I statement is true. The traditional square, unlike the modern one, allows us to infer the truth of a particular statement given the truth of its corresponding universal statement. This is because the traditional square assumes that the universal statement has existential import. It is really two statements, "All *S* is *P*" and "Some *S* exists."

Because the mood-A statement is actually two statements on the traditional interpretation, we can represent it simply by adding an additional line to our argument. It is always legitimate to change an argument by making additional assumptions. The new argument won't have the exact same impact on the audience as the old argument. The audience will now have to accept an additional premise, but in this case all we are doing is making explicit an assumption that the Aristotelian audience was making anyway. The expanded argument will look like this:

- (B1) All *M* are *P*.
- (B2) All *S* are *M*.
- (B3) Some *S* exists.\*
- (B4) ∴ Some *S* are *P*

Here the asterisk indicates that we are looking at an implicit premise that has been made explicit. This is similar to what we did in Chapter ?? on adding warrants. Now that we have an extra premise, we can add it to our Venn diagram. Since there is only one place for the *S* to be, we know where to put our *x*.

In this argument *S* is what we call the "critical term." The **CRITICAL TERM** is the term that names things that must exist in order for a conditionally valid argument to be actually valid. In this argument, the critical term was *S*, but sometimes it will be *M* or *P*.

We have used Venn diagrams to show that Barbari is valid once you include the additional premise. Using this method we can identify nine more forms, on top of the previous 15, that are valid if we add the right existence assumptions (Table 9.2)

Figure 1	Figure 2	Figure 3	Figure 4	Condition
Barbara (AAA)	Camestres (AEE)	Disamis (IAI)	Calemes (AEE)	
Celarent (EAE)	Cesare (EAE)	Bocardo (OAO)	Dimatis (IAI)	
Ferio (EIO)	Festino (EIO)	Ferison (EIO)	Fresison (EIO)	
Unconditionally Valid (Boolean Interpretation)	Baroco (Boolean Interpretation) (AOO)	Disiphi (Boolean Interpretation) (AII)		

---

Barbari (AAI)	Camestros (AEO)		Calemos (AEO)	<i>S</i> exists
Celaront (EAO)	Cesaro (EAO)			<i>S</i> exists
		Felapton (EAO)	Fesapo (EAO)	<i>M</i> exists
Conditionally Valid (Aristotelian Interpretation)	Disapti (AAI)			<i>M</i> exists
			Bamalip (AAI)	<i>P</i> exists

Table 9.2: All 24 Valid Syllogisms

Thus we now have an expanded method for evaluating arguments using Venn diagrams. To evaluate an argument, we first use a Venn diagram to determine whether it is unconditionally valid. If it is, then we are done. If it is not, then we see if adding an existence assumption can make it conditionally valid. If we can add such an assumption, add it to the list of premises and put an  $x$  in the relevant part of the Venn diagram. If we cannot make the argument valid by including additional existence assumptions, we say it is completely invalid.

Let's run through a couple examples. Consider the argument EAO-3.

- (C1) No  $M$  are  $P$ .
- (C2) All  $M$  are  $S$ .
- (C3)  $\therefore$  Some  $S$  are not  $P$ .

First we use the regular Venn diagram method to see whether the argument is unconditionally valid.

We can see from this that the argument is not valid. The conclusion says that some  $S$  are not  $P$ , but we can't tell that from this diagram. There are three possible ways something could be  $S$ , and we don't know if any of them are occupied.

Simply adding the premise  $S$  exists won't help us, because we don't know whether to put the  $x$  in the overlap between  $S$  and  $M$ , the overlap between  $S$  and  $P$ , or in the area that is just  $S$ . Of course, we would want to put it in the overlap between  $S$  and  $M$ , because that would mean that there is an  $S$  that is not  $P$ . However, we can't justify doing this simply based on the premise that  $S$  exists.

The premise that  $P$  exists will definitely not help us. The  $P$  would either go in the overlap between  $S$  and  $P$  or in the area that is only  $P$ . Neither of these would show "Some  $S$  is not  $P$ ."

The premise " $M$  exists" does the trick, however. If an  $M$  exists, it has to also be  $S$  but not  $P$ . And this is sufficient to show that some  $S$  is not  $P$ . We can then add this additional premise to the argument to make it valid.

- (D1) No  $M$  are  $P$ .
- (D2) All  $M$  are  $S$ .
- (D3)  $M$  exists.\*
- (D4)  $\therefore$  Some  $S$  are not  $P$ .

Checking it against Table 9.2, we see that we were right: this is a conditionally valid argument named Felapton.

Now consider the argument EII-3:

- (E1) No  $M$  are  $P$ .
- (E2) Some  $M$  are  $S$ .
- (E3)  $\therefore$  Some  $S$  are  $P$ .

First we need to see if it is unconditionally valid. So we draw the Venn diagram.

The conclusion says that some  $S$  are  $P$ , but we obviously don't know this from the diagram above. There is no  $x$  in the overlap between  $S$  and  $P$ . Part of that region is shaded out, but the rest could go either way.

What about conditional validity? Can we add an existence assumption that would make this valid? Well, the  $x$  we have already drawn lets us know that both  $S$  and  $M$  exist, so it won't help to add those premises. What about adding  $P$ ? That won't help either. We could add the premise " $P$  exists" but we wouldn't know whether that  $P$  is in the overlap between  $S$  and  $P$  or in the area to the right, which is just  $P$ .

Therefore this argument is invalid. And when we check the argument against Table 9.2, we see that it is not present.

### *Rules and Fallacies*

Did you do the exercises in Part ?? of Section 8? If you didn't, go back and think about those questions now, before reading any further. The problem set had five general questions like, "Can a valid argument have only negative premises?" The point of those questions was to get you to think about what patterns might exist among the 256 Aristotelian syllogisms, and how those patterns might single out 24 syllogisms as the only ones that can be valid.

In this section, we are going to answer the questions in Part ?? in detail by identifying rules that all valid syllogisms amongst the 256 Aristotelian syllogisms must obey. Seeing these rules will help you understand the *structure* of this part of logic. We aren't just assigning the labels "valid" and "invalid" to arguments randomly. Each of the rules we will identify is associated with a fallacy. If you violate the rule, you commit the fallacy.

The fallacies in this section are like the fallacies that are identified in the parts of the in Part ??in that they represent mistakes in reasoning. If you make an inference that commits one of these fallacies, you have used reasoning incorrectly. However, unlike the fallacies we looked at in the Critical Thinking section, many of these fallacies are not even tempting. They are not ways the human mind goes naturally off the rails. They are just things that you shouldn't do.

In the next subsection we are going to outline five basic rules and the fallacies that go with them, along with an addition rule/fallacy pair that can be derived from the initial five. All standard logic textbooks these days use some version of these rules, although they might divide them up differently. Some textbooks also include rules that we have built into our definition of an Aristotelian syllogism in standard form. For instance, other textbooks might have a rule here saying valid syllogisms can't have four terms, or have to use terms in the same way each time. All of this is built

into our definitions of an Aristotelian syllogism and standard form for such a syllogism, so we don't need to discuss them here.

### *Six Rules and Fallacies*

**Rule 1:** The middle term in a valid Aristotelian syllogism must be distributed at least once.

Consider these two arguments:

- |   |   |
|---|---|
| (F1) All <i>M</i> are <i>P</i> .              | (G1) All <i>P</i> are <i>M</i> .              |
| (F2) All <i>S</i> are <i>M</i> .              | (G2) All <i>S</i> are <i>M</i> .              |
| (F3) $\therefore$ All <i>S</i> are <i>P</i> . | (G3) $\therefore$ All <i>S</i> are <i>P</i> . |

The syllogism on the left (Barbara) is obviously valid, but if you change it to figure 2, you get the syllogism on the right, which is obviously invalid. What causes this change?

The premises in the second syllogism say that *S* and *P* are both parts of *M*, but they no longer tell us anything about the relationship between *S* and *P*. To see why this is the case, we need to bring back a term we saw on page 100, distribution. A term is distributed in a statement if the statement makes a claim about every member of that class. So in “All *M* are *P*” the term *M* is distributed, because the statement tells us something about every single *M*. They are all also *P*. The term *P* is not distributed in this sentence, however. We do not know anything about every single *P*. We know that *M* is in *P*, but not vice versa.

In general mood-A statements distribute the subject, but not the predicate. This means that when we reverse *P* and *M* in the first premise, we create an argument where *S* and *P* are distributed, but *M* is not. This means that the argument is always going to be invalid.

This short argument can show us that arguments with an undistributed middle are always invalid: The conclusion of an Aristotelian syllogism tries to say something about the relationship between *S* and *P*. It does this using the relationship those two terms have to the third term *M*. But if *M* is never distributed, then *S* and *P* can be different, unrelated parts of *M*. Therefore arguments with an undistributed middle are invalid. Syllogisms that violate this rule are said to commit the **FALLACY OF THE UNDISTRIBUTED MIDDLE**.

**Rule 2:** If a term is distributed in the conclusion of a valid Aristotelian syllogism, then it must also be distributed in one of the premises.

Suppose, instead of changing Barbara from a figure 1 to a figure 2 argument, we changed it to a figure 4 argument. This is what we'd get.

- (H1) All *P* are *M*.  
(H2) All *M* are *S*.



(H3)  $\therefore$  All  $S$  are  $P$ .

When we changed the argument from figure 1 to figure 2, it ceased to be valid because the middle became undistributed. But this time the middle is distributed in the second premise, and the argument still doesn't work. You can see this by filling in "animals," "mammals," and "dogs," for  $S$ ,  $M$ , and  $P$ .

(I1) All dogs are mammals.  $\Leftarrow$  True

(I2) All mammals are animals.  $\Leftarrow$  True

(I3) All animals are dogs.  $\Leftarrow$  False

This version of the argument has true premises and a false conclusion, so you know the argument form must be invalid. A valid argument form should never be able to take true premises and turn them into a false conclusion. What went wrong here?

The conclusion is a mood-A statement, which means it tries to say something about the entire subject class, namely, that it is completely contained by the predicate class. But that is not what these premises tell us. The premises tell us that the subject class, animals, is actually the broadest class of the three, containing within it the classes of mammals and dogs.

As with the previous rule, the problem here is a matter of distribution. The conclusion has the subject class distributed. It wants to say something about the entire subject class, animals. But the premises do not have "animals" as a distributed class. Premise 1 distributes the class "dogs" and premise 2 distributes the class "mammals."

Here is another argument that makes a similar mistake:

(J1) All  $M$  are  $P$ .

(J2) Some  $S$  are not  $M$ .

(J3)  $\therefore$  Some  $S$  are not  $P$ .

This time the conclusion is a mood-O statement, so the predicate term is distributed. We are trying to say something about the entire class  $P$ . But again, the premises do not say something about the entire class  $P$ .  $P$  is undistributed in the major premise.

These examples illustrate rule 2: If a term is distributed in the conclusion, it must also be distributed in the corresponding premise. Arguments that violate this rule are said to commit the **FALLACY OF ILLICIT PROCESS**. This fallacy has two versions, depending on which term is not distributed. If the subject term is the one that is not distributed, we say that the argument commits the fallacy of an illicit minor. If the predicate term isn't distributed, we say that the argument commits the fallacy of the illicit major. Some particularly silly arguments commit both.

The justification for this rule is easy enough to see. If the conclusion makes a claim about all of a class, but the premises only make a claim about

some of the class, the conclusion clearly says more than what the premises justify.

**Rule 3:** A valid Aristotelian syllogism cannot have two negative premises.

Back in exercise set ?? you were asked to determine what conclusion, if any, could be drawn from a given pair of premises. Some of the exercises involved arguments with two negative premises. Problem ?? went like this: “No  $P$  are  $M$ , and no  $M$  are  $S$ , therefore \_\_\_\_\_.” If you haven’t done so already, try to find a conclusion about  $S$  and  $P$  that you can draw from this pair of premises.

Hopefully you have convinced yourself that there is no conclusion to be drawn from the premises above using standard Aristotelian format. No matter what mood you put the conclusion is, it will not follow from the premises. The same thing would be true of any syllogism with two negative premises. We could show this conclusively by running through the 16 possible combinations of negative premises and figures. A more intuitive proof of this rule goes like this: The conclusion of an Aristotelian syllogism must tell us about the relationship between subject and predicate. But if both premises are negative then the middle term must be disjoint, either entirely or partially, from the subject and predicate terms. An argument that breaks this rule is said to commit the **FALLACY OF EXCLUSIVE PREMISES**.

**Rule 4:** A valid Aristotelian syllogism can have a negative conclusion if and only if it has exactly one negative premise.

Again, let’s start with examples, and try to see what is wrong with them.

- |  |                                     |
|--|-------------------------------------|
| (K1) All $M$ are $P$ .                   | (L1) No $P$ are $M$ .               |
| (K2) All $P$ are $M$ .                   | (L2) All $S$ are $M$ .              |
| (K3) $\therefore$ Some $S$ are not $P$ . | (L3) $\therefore$ All $S$ are $P$ . |

These arguments are so obviously invalid, you might look at them and say, “Sheesh, is there anything *right* about them?” Actually, these arguments obey all the rules we have seen so far. Look at the left hand argument. Premise 1 ensures that the middle term is distributed. The conclusion is mood O, which means the predicate is distributed, but  $P$  is also distributed in the second premise. The argument does not have two negative premises. A similar check will show that the argument on the right also obeys the first three rules.

Actually, these arguments illustrate an important premise that is independent of the previous three. You can’t draw a negative conclusion from two affirmative premises, and you cannot drawn an affirmative conclusion if there is a negative premise. Because the previous rule tells us that you can never have two negative premises, we can actually state this rule quite

simply: an argument can have a negative conclusion if and only if it has exactly one negative premise. (The phrase “if and only if” will become important when we get to SL in Chapter 10. For now you can just note that “if and only if” means that the rule goes both ways. If you have a negative conclusion, then you must have one negative premise, and if you have one negative premise, you must have a negative conclusion.)

To see why this rule is justified, you need to look at each part of it separately. First, consider the case with the affirmative conclusion. An affirmative conclusion tells us that some or all of  $S$  is contained in  $P$ . The only way to show this is if some or all of  $S$  is in  $M$ , and some or all of  $M$  is in  $P$ . You need a complete chain of inclusion. Therefore if an argument has a negative premise, it cannot have an affirmative conclusion.

On the other hand, if an argument has a negative conclusion, it is saying that  $S$  and  $P$  are at least partially separate. But if you have all affirmative premises you are never separating classes. Also, a valid argument cannot have two negative premises. Therefore, a valid argument with a negative conclusion must have exactly one negative premise.

There is not a succinct name for the fallacy that goes with violating this rule, because this is not a mistake people commonly make. We will call it the **NEGATIVE-AFFIRMATIVE FALLACY**.

**Rule 5:** A valid Aristotelian syllogism cannot have two universal premises and a particular conclusion.

This rule is a little different than the previous ones, because it really only applies if you take a Boolean approach to existential import. Consider *Barbari*, the sometimes maligned step-sister of *Barbara*:

- (M1) All  $M$  are  $P$ .
- (M2) All  $S$  are  $M$ .
- (M3)  $\therefore$  Some  $S$  are  $P$ .

This syllogism is not part of the core 15 valid syllogisms we identified with the Venn diagram method using Boolean assumptions about existential import. The premises never assert the existence of something, but the conclusion does. And this is something that is generally true under the Boolean interpretation. Universal statements never have existential import and particular statements always do. Therefore you cannot derive a particular statement from two universal statements.

Some textbooks act as if the ancient Aristotelians simply overlooked this rule. They say things like “the traditional account paid no attention to the problem of existential import” which is simply false. As we have seen, the Latin part of the Aristotelian tradition engaged in an extensive discussion of the issue from the 12th to the 16th centuries, under the heading “supposition of terms” (?). And at least some people, like William

of Ockham, had consistent theories that show why syllogisms like Barbari were valid (?).

In this textbook, we handle the existential import of universal statements by adding a premise, where appropriate, which makes the existence assumption explicit. So Barbari should look like this.

(N1) All *M* are *P*.

(N2) All *S* are *M*.

(N3) Some *S* exist.\*

(N4)  $\therefore$  Some *S* are *P*.

Adding this premise merely gives a separate line in the proof for an idea that Ockham said was already contained in premise 2. And if we make it a practice of adding existential premises to arguments like these, Rule 5 still holds true. You cannot conclude a particular statement from all universal premises. However in this case, we do have a particular premise, namely, P<sub>3</sub>. So if we provide this reasonable accommodation, we can see that syllogisms like Barbari are perfectly good members of the valid syllogism family. We will say, however, that an argument like this that does not provide the extra premise commits the **EXISTENTIAL FALLACY**.

### *Proving the Rules*

For each rule, we have presented an argument that any syllogism that breaks that rule is invalid. It turns out that the reverse is also true. If a syllogism obeys all five of these rules, it must be valid. In other words, these rules are *sufficient* to characterize validity for Aristotelian syllogisms. (For a discussion of the term “sufficient” see page [ref]). It is good practice to actually walk through a proof that these five rules are sufficient for validity. After all, that sort of proof is what formal logic is really all about. The proof below follows Hurley (<sup>2</sup> Hurley2014).

2

Imagine we have a syllogism that obeys the five rules above. We need to show that it must be valid. There are four possibilities to consider: the conclusion is either mood A, mood E, mood I, or mood O.

If the conclusion is in mood A, then we know that *S* is distributed in the conclusion. If the syllogism obeys rules 1 and 2, then we know that *S* and *M* are distributed in the premises. Rule 4 tells us that both premises must be affirmative, so the premises can’t be I or O. They can’t be E, either, because E does not distribute any terms, and we know that terms are distributed in the premises. Therefore both premises are in mood A. Furthermore, we know that they are in the first figure, because they have to distribute *S* and *M*. Therefore the syllogism is Barbara, which is valid.

Now suppose the conclusion is in mood E. By rule 4, we have one negative and one affirmative premise. Because mood-E statements distribute both subject and predicate, rules 1 and 2 tell us that all three terms must

be distributed in the premises. Therefore one premise must be E, because it will have to distribute two terms. Since E is negative, the other premise must be affirmative, and since it has to distribute a term, it can't be I. So we know one premise is A and the other E. If all the terms are distributed, this leaves us four possibilities: EAE-1, EAE-2, AEE-2, and AEE-4. These are the valid syllogisms Celarent, Cesare, Camestres, and Calemes.

Next up, consider the case where the conclusion is in mood I. By rule 4, it has two affirmative premises, and by rule 5 both premises cannot be universal. This means that one premise must be an affirmative particular statement, that is, mood I. But we also know that by rule 1 some premise must distribute the middle term. Since this can't be the mood-I premise, it must be the other premise, which then must be in mood A. Again we are reduced to four possibilities: AII-1, AII-2, IAI-3, and IAI-4, which are the valid syllogisms Darii, Datisi, Disamis, and Dimatis.

Finally, we need to consider the case where the conclusion is mood O. Rule 4 tells us that one premise must be negative and the other affirmative, and rule 5 tells us that they can't both be universal. Rules 1 and 2 tell us that *M* and *P* are distributed in the premises. This means that the premises can't both be particular, because then one would be I and one would be O, and only one term could be distributed. So one premise must be negative and the other affirmative, and one premise must be particular and the other universal. In other words, our premises must be a pair that goes across the diagonal of the square of opposition, either an A and an O or an E and an I.

With the AO pair, there are two possibilities that distribute the right terms: OAO-3 and AOO-II. These are the valid syllogisms Bocardo and Baroco. With the EI pair, there are four possibilities, which are all valid. They are the EIO sisters: Ferio, Festino, Ferison, and Fresison.

So there you have it. Those five rules completely characterize the possible valid Aristotelian syllogisms. Any other patterns you might notice among the valid syllogisms can be derived from these five rules. For instance, Problem ?? in exercise set ?? of Section 8 asked if you could have a valid Aristotelian syllogism with two particular premises. If you did that problem, hopefully you saw that the answer was "no." We could, in fact, make this one of our five rules above. But we don't need to. When we showed that these five rules were sufficient to characterize validity, we also showed that any other rule characterizing validity that we care to come up with can be derived from the rules we already set out. So, let's state the idea that a syllogism cannot have two particular premises as a rule, and show how it can be derived. This will be our statement of the rule:

**Derived Rule 1:**      A valid Aristotelian syllogism cannot have two particular premises.

And let's call the associated fallacy the **FALLACY OF PARTICULAR PREMISES**. To show that this rule can be derived from the previous five,

it is sufficient to show that any syllogism that violates this rule will also violate one of the previous five rules. Thus there will always be a reason, independent of this rule, that can explain why that syllogism is false.

So suppose we have a syllogism with two particular premises. If we want to avoid violating rule 1, we need to distribute the middle term, which means that both premises cannot be mood I, because mood-I statements don't distribute any term. We also know that both statements can't be mood O, because rule 3 says we can't have two negative premises. Therefore our syllogism has one premise that is I and one premise that is O. It thus has exactly one negative premise, and by rule 4, must have a negative conclusion, either an E or an O. But an argument with premises I and O can only have one term distributed: if the conclusion is mood O, then two terms are distributed; and if it is mood E then all three terms are distributed. Thus any syllogism that manages to avoid rules 1, 3, and 4 will fall victim to rule 2. Therefore any syllogism with two particular premises will violate one of the five basic rules.

### *Validity and the Counterexample Method*

Except for a brief discussion of logically structured English in section 8, so far, we have only been evaluating arguments that use variables for the subject, middle, and predicate terms. Now, we will be looking in detail at issues that come up when we try to evaluate categorical arguments that come up in ordinary English. In this section we will consider how your knowledge of the real world terms discussed in an argument can distract you from evaluating the form of the argument itself. In this section we will consider difficulties in putting ordinary language arguments into logically structured English, so they can be analyzed using the techniques we have learned.

Let's go back again to the definition of validity on page ?? (It is always good for beginning student to reinforce their understanding of validity). We did this in the last chapter in Section ??, and we are doing it again now. Validity is a fundamental concept in logic that can be confusing. A valid argument is not necessarily one with true premises or a true conclusion. An argument is valid if the premises *would* make the conclusion true *if* the premises were true.

This means that, as we have seen before, there can be valid arguments with false conclusions. Consider this argument:

No reptiles are chihuahuas. But all dogs are reptiles. Therefore, no dogs are chihuahuas.

This seems silly, if only because the conclusion is false. We know that some dogs are chihuahuas. But the argument is still valid. In fact, it shares a form with an argument that makes perfect sense:

No reptiles are dogs, but all chameleons are reptiles. Therefore, no dogs are chameleons.

Both of these arguments have the form Celarent:

- (O1) No  $M$  are  $P$ .
- (O2) All  $S$  are  $M$ .
- (O3)  $\therefore$  No  $S$  are  $P$ .

This form is valid, whether the subject and predicate term are dogs and chameleons, or dogs and chihuahuas, which you can see from this Venn diagram.

This means you can't assume an argument is invalid because it has a false conclusion. The reverse is also true. You can't assume an argument is valid just because it has a true conclusion. Consider this

All cats are animals, and some animals are dogs. Therefore no dogs are cats.

Makes sense, right? Everything is true. But the argument isn't valid. The premises aren't making the conclusion true. Other arguments with the same form have true premises and a false conclusion. Like this one.

All chihuahuas are animals, and some animals are dogs. Therefore, no dogs are chihuahuas.

Really, the arguments in both these passages have the same form: AEE-IV:

- (P1) All  $P$  are  $M$ .
- (P2) Some  $M$  are  $S$ .
- (P3)  $\therefore$  No  $S$  are  $P$ .

This is an invalid form, and it remains invalid whether the  $P$  stands for cats or chihuahuas. You can see this in the Venn diagram:

All these examples bring out an important fact about the kind of logic we are doing in this chapter and the last one: this is *formal* logic. As we discussed on page ?? formal logic is a way of making our investigation content neutral. By using variables for terms instead of noun phrases in English we can show that certain ways of arguing are good or bad regardless of the topic being argued about. This method will be extended in Chapters 10 and ??, when we introduce our full formal languages SL and QL. This method will be extended in Part ??, when we introduce the full formal language SL.

The examples above also show us another way of proving that an argument given in English is invalid, called the counterexample method. As we have just seen, if you are given an argument in English with, say, false premises and a false conclusion, you cannot determine immediately

whether the argument is valid. However, we can look at arguments that have the same form, and use them to see whether the argument is valid. If we can find an argument that has the exact same form as a given argument but has true premises and a false conclusion, then we know the argument is invalid. We just did that with the AEE-IV argument above. We were given an argument with true premises and a true conclusion involving cats, dogs, and animals. We were able to show this argument invalid by finding an argument with the same form that has true premises and a false conclusion, this time involving chihuahuas, dogs, and animals.

More precisely, we can define the **COUNTEREXAMPLE METHOD** as a method for determining if an argument with ordinary English words for terms is valid, where one consistently substitutes other English terms for the terms in the given argument to see if one can find an argument with the same form that has true premises and a false conclusion. Let's run through a couple more examples to see how this works.

First consider this argument in English:

All tablet computers are computers. We know this because a computer is a kind of machine, and some machines are not tablet computers.

Every statement in this argument is true, but it doesn't seem right. The premises don't really relate to the conclusion. That means you can probably find an argument with the same form that has true premises and a false conclusion. Let's start by putting the argument in canonical form. Notice that the English passage had the conclusion first.

(Q1) All computers are machines.

(Q2) Some machines are not tablet computers.

(Q3)  $\therefore$  All tablet computers are computers.

Let's find substitutes for "machines," "computers," and "tablet computers" that will give us true premises and a false conclusion. It is easiest to work with really common sense categories, like "dog" and "cat." It is also easiest to start with a false conclusion and then try to find a middle term that will give you true premises. "All dogs are cats" is a nice false conclusion to start with:

(R1) All cats are  $M$ .

(R2) Some  $M$  are not dogs.

(R3)  $\therefore$  All dogs are cats.

So what can we substitute for  $M$  (which used to be "machines") that will make  $P_1$  and  $P_2$  true? "Animals" works fine.

(S1) All cats are animals.

(S2) Some animals are not dogs.



(S3)  $\therefore$  All dogs are cats.

There you have it: a counterexample that shows the argument invalid. Let's try another one.

Some diseases are viral, therefore some things caused by bacteria are not things that are caused by viruses, because all diseases are bacterial.

This will take a bit more unpacking. You can see from the indicator words that the conclusion is in the middle. We also have to fix "viral" and "things that are caused by viruses" so they match, and the same is true for "bacterial" and "things that are caused by bacteria." Once we have the sentences in canonical form, the argument will look like this:

(T1) Some diseases are things caused by viruses.

(T2) All diseases are things caused by bacteria.

(T3)  $\therefore$  Some things caused by bacteria are not things caused by viruses.

Once you manage to think through the complicated wording here, you can see that  $P_1$  and the conclusion are true. Some diseases come from viruses, and not everything that comes from a bacteria comes from a virus. But  $P_2$  is false. All diseases are not caused by bacteria. In fact,  $P_1$  contradicts  $P_2$ . But none of this is enough to show the argument is invalid. To do that, we need to find an argument with the same form that has true premises and a false conclusion.

Let's go back to the simple categories: "dogs," "animals," etc. We need a false conclusion. Let's go with "Some dogs are not animals."

(U1) Some  $M$  are dogs.

(U2) All  $M$  are animals.

(U3)  $\therefore$  Some dogs are not animals.

We need a middle term that will make the premises true. It needs to be a class that is more general than "dogs" but more narrow than "animals." "Mammals" is a good standby here.

(V1) Some mammals are dogs.

(V2) All mammals are animals.

(V3)  $\therefore$  Some dogs are not animals.

And again we have it, a counterexample to the given syllogism.

### *Ordinary Language Arguments*

In the last section we saw that arguments in ordinary language can throw us off because our knowledge of the truth or falsity of the statements in them can cloud our perception of the validity of the argument. Now we are going to look at ways we might need to transform ordinary language arguments in order to put them in standard form (see p. ??) and use our evaluative tools.

### *Basic Transformations into Logically Structured English*

Arguments are composed of statements, so all of the guidelines we discussed for putting ordinary English statements into logically structured English in Section 7 apply here. Also, as you saw in the exercises in Section 8, we need to be aware that the conclusion of an ordinary language argument can occur anywhere in the passage. Consider this argument

Every single penguin in the world is a bird. Therefore, at least one bird in the world is not a person, because not one person is a penguin.

The indicator words let us know the conclusion is the sentence in the middle of this passage, “At least one bird in the world is not a person.” The quantifier in this sentence is nonstandard. As we saw on page ?? “at least one” should be changed to “some.” That would give us “Some birds are not people” for the conclusion. Similar changes need to be made to the two premises. Thus the argument and translation key look like this:

S: Birds	(W1) No <i>P</i> are <i>M</i> .
M: Penguins	(W2) All <i>M</i> are <i>S</i> .
P: People	(W3) ∴ Some <i>S</i> are not <i>P</i> .

Now that we have it in this form, we can see that it is the argument Fesapo (EAO-4), which is valid if you add the premise that some penguins exist.

Another issue that can come up in ordinary language arguments is the use of synonyms. Consider another argument about penguins:

No Adélie penguins can fly, because all members of *Pygoscelis adeliae* are members of the family Sphenisciformes, and no Sphenisciformes can fly.

You might think this argument can't be a proper syllogism, because it has four terms in it: Adélie penguins, *Pygoscelis adeliae*, Sphenisciformes, and things that can fly. However, a quick trip to Wikipedia can confirm that *Pygoscelis adeliae* is just the scientific name from the Adélie penguin, and for that matter “Sphenisciformes” are just penguins. So really, the argument just has three terms, and there is no problem representing the argument like this:

S: Adélie penguins	(X1) No <i>M</i> are <i>P</i> .
M: Penguins	(X2) All <i>S</i> are <i>M</i> .
P: Things that can fly	(X3) ∴ No <i>S</i> are <i>P</i> .

And this is our good friend Celarent (EAE-1).

Generally speaking, arguments involving different names for the same person can work the same way as synonyms. Consider this argument:

Bertrand Russell was a philosopher. We know this because Mr. Russell was a logician, and all logicians are philosophers.

“Bertrand Russell” and “Mr. Russell” are names for the same person, so we can represent them using the same variable. Back on page ?? we learned that names for individuals need to be transformed into classes using phrases like “objects identical with ...” or “people identical with ...” Thus the argument turns out to be Celarent’s best friend, Barbara (AAA-I):

S: People identical with Bertrand Russell	(Y1) All <i>M</i> are <i>P</i> .
	(Y2) All <i>S</i> are <i>M</i> .
M: Logicians	(Y3) ∴ All <i>S</i> are <i>P</i> .
P: Philosophers	

Merging proper nouns into one variable doesn’t always work the way merging synonyms does, however. Sometimes exchanging one name with another can change the truth value of a sentence, even when the two names refer to the same person. Any comic book fan will tell you that the sentence “J. Jonah Jameson hates Spider-Man” is true. Also, “Peter Parker” and “Spider-Man” refer to the same person. However “J. Jonah Jameson hates Peter Parker” is false. These situations are called “intentional contexts.” We won’t be dealing with them in this textbook, but you should be aware that they exist.

Another thing to be aware of when you assign variables to terms is that often you will need to find the right general category that will allow your terms to hook up with each other. Consider this argument.

Bertrand Russell did not believe in God. I know this because Mr. Russell was not a Christian, and all Christians believe in God

Here again we need to use one variable for both “Bertrand Russell” and “Mr. Russell.” But we also have to change the verb phrase “believes in God” into a predicate term (see page ??). The main trick here is to transform the verb phrase “believes in God” and the singular terms referring to Bertrand Russell in a way that will match. The trick is to use the word “people” each time. “Believes in God” becomes “people who believe in God,” and the two names for Bertrand Russell need to become “people identical with Bertrand Russell.” By including the word “people” in each case, we ensure that the terms will be comparable.

S: People identical to Bertrand Russell	(Z1) All <i>M</i> are <i>P</i> .
	(Z2) No <i>S</i> are <i>M</i> .
M: Christians	(Z3) ∴ No <i>S</i> are <i>P</i> .
P: People who believe in God	

The argument here is AEE-1, which is invalid. It violates rule 2, because  $P$  is distributed in the conclusion, but not the premises. More informally, we can say that just because Russell wasn't a Christian doesn't mean he was an atheist or an agnostic. (In fact, Russell said that he was an atheist in some ways, and an agnostic in others, but this argument doesn't show that.)

### *Complement Classes, Obversion, and Contraposition*

In the last subsection, we saw that we can represent synonyms using the same variable, thus reducing the number of terms in an argument down to the three terms required for an Aristotelian syllogism. Something similar can be done when you have terms that are complements of other terms. Back on page ?? we said that the complement of a class was the class of things that are not in the original class. So the complement of the class "penguins" is everything that is not a penguin. Sometimes, if you have an argument with more than three terms, you can reduce the number of terms down to three, because some of the terms are complements of each other, like "penguins" and "non-penguins."

Consider the following argument, which talks about the stoics, an ancient philosophical school and pioneers in logical thinking.

No stoics are non-logicians, and all non-philosophers are non-stoics. Therefore all logicians are philosophers.

This argument has six terms: stoics, logicians, philosophers, non-stoics, non-logicians, and non-philosophers. Since the second three are all complements of the first, we should be able to reduce the number of terms to three. Once we do this, we might find that the argument is valid—it looks like it might be Barbara. We can begin by assigning  $S$ ,  $M$ , and  $P$  to the terms "stoics," "logicians," and "philosophers." We can then represent the other six terms as non- $S$ , non- $M$ , and non- $P$ .

$S$ : Logicians	(AA1) No $M$ are non- $P$ .
$M$ : Stoics	(AA2) All non- $S$ are non- $M$ .
$P$ : Philosophers	(AA3) $\therefore$ All $S$ are $P$ .

The argument still has six terms, so we can't evaluate it as is. However, we can get rid of three of these terms. The secret is to apply two of the transformation tools we learned about in Section ??, obversion and contraposition. In that section, we saw that there were different ways you could alter categorical statements that would sometimes yield a logically equivalent statement. Two of those methods—obversion and contraposition—involved taking the complements of existing terms. In the cases where these transformations yield logically equivalent terms, we can use them to change the

premises of an argument into statements that lack the terms we want to get rid of.

In obversion, you change the quality of a sentence—switch it from affirmative to negative or vice versa—and then replace the predicate term with its complement. This process always yields a logically equivalent statement, so we can always use it. Applying obversion to the first premise of the argument above allows us to eliminate the term *non-P*.

(AB1) All *M* are *P*.

(AB2) All *non-S* are *non-M*.

(AB3) ∴ All *S* are *P*.

We can also use contraposition to eliminate terms, but here we must be more careful. Contraposition only yields logically equivalent statements when it is applied to mood-A or mood-O statements. Fortunately, Premise 2 is a mood-A statement, so we can change the argument to look like this:

(AC1) All *M* are *P*.

(AC2) All *M* are *S*.

(AC3) ∴ All *S* are *P*.

And now the argument only has three terms, and we can evaluate it. As you can see it is not Barbara at all, but one of her unnamed evil step-sisters, AAA-3. Again, remember not to use contraposition on mood-E or mood-I statements. You can use obversion on any statement, but contraposition only works on mood A or mood O.

Ordinary English has other prefixes besides “non-” to denote the complements of classes. “Un-” is the most common prefix, but English also uses “in-,” “dis-,” and “a-.” These can be handled the same way “non-” is. Consider,

All dogs in the park must be leashed. Some unleashed pets in the park are cats. Therefore, some cats in the park are not dogs.

To make this argument work, we need to get the predicate of the major premise to match the subject of the minor premise. The first step is just turning the verb phrase “must be leashed” into a noun phrase that matches “unleashed pets.” We also should use the qualification “in the park” consistently across the terms.

*S*: Cats in the park

(AD1) All *P* are *M*.

*M*: Leashed pets in the park

(AD2) Some *non-M* are *S*.

(AD3) ∴ Some *S* are not *P*.

*P*: Dogs in the park

Now the trick is to get the terms *M* and *non-M* to match. We could try to transform the second premise, but that won’t do us any good. It is an E

statement, so we can't use contraposition, and obversion won't change the subject term.

Instead of changing the "non- $M$ " to " $M$ " in the second premise, we need to change the " $M$ " to "non- $M$ " in the first premise. We can do this using obversion, which works on statements in any mood.

(AE1) No  $P$  are non- $M$ .

(AE2) Some non- $M$  are  $S$ .

(AE3)  $\therefore$  Some  $S$  are not  $P$ .

And now we can see that the argument is Fesison (EIO-4), which is valid.

### *Enthymemes*

In the real world, arguments are often presented with pieces missing. The missing piece could be an important premise or sometimes even the conclusion. Arguments with parts missing are called **ENTHYMEMES**. We discuss them extensively in the chapter on incomplete arguments in the full version of this text. Here we will be dealing with them specifically as they are used with Aristotelian syllogisms. Because we are now dealing with more real-world arguments, we will return to our practice of giving the context in italics before passages (see page 28). Remember, that this context is not a part of the argument. It is just there to help interpret the passage.

The simplest and most common reason one might leave out a part of an Aristotelian syllogism is brevity. Sometimes part of an argument is common knowledge or completely obvious from the context and simply doesn't need to be said. Consider this common inference that might pop into your head one afternoon:

*Susan is working from home* Well, the dog is barking. I guess that means the mail is here.

Here the missing premise is that dogs bark at letter carriers. But this isn't something Susan is going to think of consciously. It is completely common knowledge, and probably something Susan has a lot of first-hand experience with as a dog owner. So the passage above basically represents Susan's inference as it occurs to her.

If in order to evaluate the argument, though, we need to make all of the reasoning explicit. This means putting the whole thing in standard form and then finding the missing premise. Once you do that, it becomes clear that this common, everyday inference is actually very complicated. The simplest way to put the argument in canonical form would simply be this:

(AF1) The dog is barking.

(AF2)  $\therefore$  The mail is here.

But if we are going to use the tools of Aristotelian logic to show that this is valid, we are going to need to convert these statements into categorical statements. The “now” functions here as a singular term, so we needed to use the awkward “times identical with” construction (see page ??). Also, we needed to find a general category to put things under, in this case “times,” to be sure that all the terms matched. (This is similar to the “persons identical to Bertrand Russell example on page 139.)

(AG1) All times identical with now are times when the dog is barking.

(AG2) ∴ All times identical with now are times when the mail arrives.

Once we do this, we can see that the conclusion is a statement in mood A. The minor term is “times identical with now” and the major term is “times when the mail arrives.” The premise is a minor premise, because it has the minor term in it, and the middle term is “times when the dog is barking.” From this we can see that the missing premise is the major premise. When we add it, the full argument looks like this:

(AH1) All times when the dog is barking are times when the mail arrives\*

(AH2) All times identical with now are times when the dog is barking.

(AH3) ∴ All times identical with now are times when the mail arrives.

Remember that the asterisks means that an implicit premise has been made explicit. (See Chapter ??) Once we have that in place, however, we can see that this argument is just a version of Barbara (AAA-1). The full argument is quite wordy, which shows how much sophisticated reasoning is going on in your brain when you make even the most everyday of inferences.

Enthymemes can also be used to hide premises that are needed for the argument, but that the audience might not accept if their attention were drawn to them. Consider this argument:

*A used car salesperson is showing you a Porsche Cayenne Hybrid SUV And it is a hybrid electric car, so you know it gets good gas mileage.*

Here again we have an argument about a singular object—this car the salesperson is showing you. So when we represent the argument, we need to use the awkward “things identical with” construction.

(AI1) All things identical with this vehicle are hybrid electric vehicles.

(AI2) ∴ All things identical with this vehicle are vehicles that get good gas mileage.

Again, we have two mood-A statements. The minor term is “things identical with this vehicle,” the middle term is “hybrid electric vehicles” and the major term is “vehicles that get good gas mileage.” The missing premise must be the major premise “All hybrid electric vehicles get good gas mileage.”

- (A71) All hybrid electric vehicles are vehicles that get good gas mileage.\*  
 (A72) All things identical with this vehicle are hybrid electric vehicles.  
 (A73)  $\therefore$  All things identical with this vehicle are vehicles that get good gas mileage.

But wait, is this implicit premise even true? Actually it isn't—the Cayenne itself only gets 20 miles per gallon in the city and 24 on the highway—no wonder the salesperson kept this premise implicit! Compare the Cayenne's mileage to that of a regular gas-powered car like the Honda Fit, which gets 33 city and 41 highway. It is true that hybrids generally get better mileage than similarly sized cars that run on gasoline only, but that isn't the reasoning the salesperson is presenting here. Someone who was really all that concerned about mileage probably should not be in the market for a luxury SUV like the Cayenne.

When filling in missing parts of an enthymeme it is important to remember the principle of charity. The principle of charity says that we should try to interpret other people in a way that makes reasoning coherent and their beliefs reasonable. (See Chapter ??.) With the example of the salesperson above, one might argue that we were not being as charitable as we could have been. Perhaps the salesperson didn't expect us to believe that *all* hybrids get good gas mileage. Perhaps they only expected us to believe that *most* hybrids get good gas mileage.

If you interpret the argument this way, you are getting a more believable premise in exchange for a weaker inference. The original argument had the strongest possible inference—it was genuinely valid. If you change the major premise to something involving “most,” you will have an argument that is at best strong, not valid. The premise is now believable, but the inference is fallible. (This kind of trade-off is discussed more extensively in Chapter ??) Also note that if you do decide to use “most” when filling in the missing premise, the argument becomes inductive, and thus is something that you would evaluate using the tools from the parts of this textbook on inductive and scientific reasoning, not the tools of Aristotelian logic. In this section we will only be looking at enthymemes where you fill in premises in a way that yields Aristotelian syllogisms.

Sometimes enthymemes are missing a conclusion, rather than a premise. Consider this example:

*Annabeth, age 7, is pestering her mother Susan with strange questions.*

**Annabeth** Mommy, do any dogs have forked tongues?

**Susan:** Well, dear, all dogs are mammals, and no mammals have forked tongues. So do you think any dogs have forked tongues?

Teachers and parents often replace the conclusion with a rhetorical question like this because it forces the audience to fill in the missing con-



clusion for themselves. The effect is going to be different for different audiences, however. A child might be pleased with herself for figuring out the answer with just a few hints. Often replacing the conclusion with a rhetorical question when talking to an adult can be demeaning, for instance, when someone glibly says “you do the math,” even when there is no math involved in the issue.

In any case, the proper conclusion for the argument above is “No dogs have forked tongues.” The argument is Celarent (EAE-1), and in canonical form it looks like this:

(AK1) No mammals have forked tongues.

(AK2) All dogs are mammals.

(AK3)  $\therefore$  No dogs have forked tongues.

Evaluating an enthymeme requires four steps. First, you need to figure out if the missing statement is a premise or a conclusion. As you might expect, indicator phrases will be helpful here. Words like “because” or “therefore” will mark one of the sentences you are given as a premise or a conclusion, and you can use this information to determine what hasn’t been given. In fact, if the enthymeme contains any indicator words at all, you can be pretty confident that the missing statement is a premise. If the enthymeme contains a conclusion indicator word like “therefore,” you know that the statement after it is the conclusion. But even if the indicator word is a premise indicator word, you the missing statement is still likely to be a premise, because the premise indicator words are generally preceded by conclusions. Consider this following argument.

*Susan is now wondering about what reptilian characteristics some mammals do have* Well, I know some mammals have scales, because armadillos have scales.

Here the “because” lets us know that “Armadillos have scales” is a premise, and “some mammals have scales” is the conclusion. Thus the missing statement is one of the premises.

Once you have figured out what part of the syllogism is missing, the second step is to write the parts of the syllogism you do have in standard form. For this example, that is pretty easy, but you do need to provide the missing quantifier for the premise.

(AL1) All armadillos have scales.

(AL2)  $\therefore$  Some mammals have scales.

Step three is to figure out what terms are in the missing statements. The two statements you have been given will contain a total of three terms, one of which will be repeated once in each statement. Since every term in an Aristotelian syllogism appears twice, we know that the two terms that only appear once must be the ones that appear in the missing statement. In the

example above, the terms are “mammals,” “armadillos,” and “things with scales.” The term “things with scales” is the one that appears twice. So the missing statement will use the terms “armadillos” and “mammals.”

Step four is to determine the mood of the missing statement and the overall figure of the syllogism. This will let you fill in everything else you need to know to complete the syllogism. The rules for a valid syllogism listed in Section 9 can be a big help here.

In the example we have been working with we know that the conclusion is “Some mammals have scales.” This means that the major term is “things with scales” and the minor term is “mammals.” The premise that we are given, “All armadillos have scales” must then be the major premise. So we know this much:

(AM1) All armadillos have scales.

(AM2) [something with armadillos and mammals]

(AM3)  $\therefore$  Some mammals have scales.

We also know that the middle term is in the subject position of the major premise, which means that the whole syllogism is either figure 1 or figure 3. Because the major premise and the conclusion are both affirmative, we know by Rule 4 (page 130) that the minor premise must be affirmative. That leaves us only mood-A and mood-I statements.

So there are four possibilities remaining: the argument is either AAI-1, AII-1, AAI-3 or AII-3. A quick glance at the Table 9.2 tells us that all of these are valid. (Although AAI-1 and AAI-3 are conditionally valid.) So any of these are correct answers. Let’s go with AII-1 (Dari):

(AN1) All armadillos have scales

(AN2) Some mammals are armadillos.

(AN3)  $\therefore$  Some mammals have scales.

Sometimes, when people give enthymemes, there is no way to fill in the missing premise that will make the argument valid. This can happen either because the person making the argument is confused or because they are being deceptive. Let’s look at one example:

*Now Annabeth wants a pet armadillo* If an animal makes a good pet, then it must be cute. So armadillos must make good pets.

There is no way to fill in Annabeth’s reasoning here that can turn this into a valid syllogism. To see this, we need to first put it in standard form, which means converting the conditional “if...then” statement into a mood-A statement.

(AO1) All animals that make good pets are animals that are cute.

(AO2)  $\therefore$  All armadillos are animals that make good pets.

It seems like the missing premise here should be “All armadillos are cute,” but adding that premise doesn’t give us a valid argument.

(AP1) All animals that make good pets are animals that are cute.

(AP2) All armadillos are animals that are cute.

(AP3)  $\therefore$  All armadillos are animals that make good pets.

The argument is AAA-2, which has an undistributed middle. In fact, there is no way to get from  $P_1$  to the conclusion with an Aristotelian categorical statement. The major term has the premise in the subject position, which means that the argument can only be figure 2 or 4. But a quick look at Table 9.2 lets us know that the only valid argument with a mood-A major premise and a mood-A conclusion is Barbara, which is figure 1. Nice try, kid.

In other cases, the rules given in Section 9 can be used to determine whether an enthymeme with a missing conclusion can be valid. In the argument showing that some mammals have scales (p. 146), we used Rule 4 to show that the missing premise had to be affirmative. In an enthymeme with a missing conclusion, you might note that there are two negative premises, so that there actually is no valid conclusion to draw. In other situations you might be able to determine that the missing conclusion must be negative. The rules for valid syllogisms are your friends when working with enthymemes, because they allow you to consider broad categories of syllogisms, rather than having to work on a trial and error basis with the Venn diagram method.

### *Sorites Categorical Arguments*

The Aristotelian tradition mostly focused on syllogisms with two premises; however, there are fun things we can do with longer arguments. These were explored, among other places, in a 19th century textbook called *Symbolic Logic* (?), by the mathematician and logician Charles Lutwidge Dodgson. (Dodgson is actually better known by his pen name, Lewis Carroll, under which he wrote the children’s classics *Alice’s Adventures in Wonderland* and *Through the Looking Glass*.)

Categorical arguments with more than two premises are called **SORITES CATEGORICAL ARGUMENTS**, or just “Sorites” (pronounced “sore-EYE-tease”) for short. The name comes from the Greek word “Soros” meaning “heap.” This kind of sorites should not be confused with the sorites paradoxes we talk about in the chapter on real world evaluation in the complete version of this text, Chapter ??, which were arguments that exploited vague premises.

Here is a simple example of a categorical sorites.

(AQ1) All carrots are vegetables.

(AQ2) No vegetables are houses.

(AQ3) All houses are buildings.

(AQ4)  $\therefore$  No carrots are buildings.

We have more premises and terms than we do in the classic Aristotelian syllogism, but they are arranged in a way that is a natural extension of the Aristotelian syllogism. Every term appears twice, except for the major and minor terms of the conclusion. Each premise shares one term with each of the other premises. Really, this is just like a regular Aristotelian syllogism with more middle terms.

Because the form of the argument above is an extension of the standard form of an Aristotelian syllogism, you can actually break it apart into two valid Aristotelian syllogisms if you supply the missing statements.

(AR1) All carrots are vegetables.

(AR2) No vegetables are houses.

(AR3)  $\therefore$  No carrots are houses.\*

(AS1) No carrots are houses.\*

(AS2) All houses are buildings.

(AS3)  $\therefore$  No carrots are buildings.

The argument on the left takes the first two premises from the sorites argument and derives a conclusion from them that was not stated in the original argument. We then slide this conclusion over and use it as the major premise for the next argument, where it gets combined with the final premise and the conclusion of the sorites argument. The result is two valid arguments, Calemes and Celarent, which have been chained together. The conclusion of the Calemes argument is an intermediate conclusion, and the conclusion of the Celarent is the final conclusion. You can think of the sorites argument as a sort of abbreviation for this chain of valid arguments.

The task for this section is to evaluate the validity of sorites arguments. To do this, we will need to put them in standard form, just as we did for ordinary Aristotelian arguments. We will define **STANDARD FORM FOR A SORITES CATEGORICAL ARGUMENT** as a sorites argument that has been put into logically structured English with the following criteria: (1) each statement in the argument is in standard form for a categorical statement in logically structured English, (2) each instance of a term is in the same format and is used in the same sense, (3) the major premise is first in the list of premises and the minor premise is last, and (4) the middle premises are arranged so that premises that share a term are adjacent to one another.

Some passages will take more work to get into standard form for sorites than others. We will just look at situations where you need to rearrange the order of the premises and fix the terms so that they match. We will also have an example using variables for terms. We will use the letters *A*, *B*, *C*, ... for the terms in sorites categorical arguments, rather than *S*, *M*, and *P*.

(AT1) All *D* are *E*.

(AT2) Some *C* are *D*.

(AT3) All *C* are *B*.

(AT4) All *A* are non-*B*.

(AT5) ∴ Some *E* are not *A*.

In this argument, *P*<sub>3</sub> and *P*<sub>4</sub> don't match up, because one talks about *B* and the other non-*B*. We can just use obversion on *P*<sub>4</sub>:

(AU1) All *D* are *E*.

(AU2) Some *C* are *D*.

(AU3) All *C* are *B*.

(AU4) No *A* are *B*.

(AU5) ∴ Some *E* are not *A*.

Now we need to put the premises in the proper order. The predicate for the conclusion is *A*, so the statement “No *A* are *B*” needs to be the first premise. The statement containing *E* needs to be the last premise, and the middle two premises need to be swapped so each premise will share a term with the statements on either side of it.

(AV1) No *A* are *B*.

(AV2) All *C* are *B*.

(AV3) Some *C* are *D*.

(AV4) All *D* are *E*.

(AV5) ∴ Some *E* are not *A*.

In this example, the letters wind up in ordinary alphabetical order. Not every example will work this way.

Once we have the arguments in standard form, our job is to evaluate them. We will look at three methods for doing this. The first will require you to break down the sorites into its component syllogisms and evaluate them separately. The second two will let you evaluate the argument all at once.

### *Checking Each Inference with Venn Diagrams*

The most thorough way to check a categorical sorites argument is to break it down into its component Aristotelian syllogisms and check each of those separately. Look at the argument on the previous page. We need to identify the implicit intermediate conclusions and write out each syllogism separately. The first two premises of this argument are “No *A* are *B*,” and “All *C* are *B*.” That's a mood-E statement and a mood-A statement, with both middle terms on the right. A glance at Table 9.2 lets us know that we can conclude a mood-E statement from this, “No *C* are *A*.” So the first component argument is Cesare (EAE-II).

(AW1) No *A* are *B*.

(AW2) All *C* are *B*.

(AW3) ∴ No *C* are *A*.\*

The conclusion of this argument, “No  $C$  are  $A$ ,” is an implicit intermediate step, which means it is the major premise for the next component syllogism. The minor premise is  $P_3$ , “Some  $C$  are  $D$ .” This gives us a mood-E and a mood-O statement. The new middle term is  $C$ . Again we can consult Table 9.2 to see that this sets up Ferison (EIO-3):

- (AX1) No  $C$  are  $A$ .  
 (AX2) Some  $C$  are  $D$ .  
 (AX3)  $\therefore$  Some  $D$  are not  $A$ .

The conclusion here becomes the major premise of the next component argument. The minor premise is  $P_4$  of the original sorites. This means that the last step is Bocardo (OAO-3).

- (AY1) Some  $D$  are not  $A$ .  
 (AY2) All  $D$  are  $E$ .  
 (AY3)  $\therefore$  Some  $E$  are not  $A$ .

To wrap things up, we can use three Venn diagrams to confirm that each step is valid. This step is needed to be sure that the terms appear in the correct position in each diagram. The three Venn diagrams are shown together with the corresponding arguments in Figure ?? . Notice that the major term is always represented by the lower right circle: it is the predicate of the conclusion of each component argument. The subject of the conclusion for each component argument then becomes the middle term for the next argument.

This example turned out to be valid, but not every sorites you will be asked to evaluate will work that way. Consider this example, given in standard form.

- (AZ1) All  $B$  are  $A$ .  
 (AZ2) All  $C$  are  $B$ .  
 (AZ3) Some  $C$  are not  $D$ .  
 (AZ4)  $\therefore$  No  $D$  are  $A$ .

If you look at the first two premises, you can easily see that they entail “All  $C$  are  $A$ .” The first component argument is a simple Barbara. But if you plug that conclusion into the next component argument, the result is invalid.

- |                                      |                                     |
|--------------------------------------|-------------------------------------|
| (BA1) All $B$ are $A$ .              | (BB1) All $C$ are $A$ .             |
| (BA2) All $C$ are $B$ .              | (BB2) Some $C$ are not $D$ .        |
| (BA3) $\therefore$ All $C$ are $A$ . | (BB3) $\therefore$ No $D$ are $A$ . |
| <b>Barbara (AAA-I)</b>               | <b>AOE-III (Invalid)</b>            |

You will also encounter situations where two early premises in a sorites don't lead to any valid conclusion, so there is simply no way to fill in the missing steps. Consider this:

- (BC1) No *D* are *B*.  
 (BC2) No *D* are *C*.  
 (BC3) All *A* are *C*.  
 (BC4) ∴ All *A* are *B*.

Here the first two premises are negative, so we know by Rule 3 that there is no conclusion we can validly draw here. Thus there is nothing to plug in as the major premise for the second component argument.

(BD1) No <i>D</i> are <i>B</i> .	(BE1) ???
(BD2) No <i>D</i> are <i>C</i> .	(BE2) All <i>A</i> are <i>C</i> .
(BD3) ∴ ???	(BE3) ∴ All <i>A</i> are <i>B</i> .
<b>Incomplete</b>	<b>Incomplete</b>

Actually, once you note that the first two premises violate Rule 3 for Aristotelian syllogisms, there really isn't a need to break down the argument further. This brings us to the second method for evaluating sorites categorical arguments.

### *Checking the Whole Sorites Using Rules*

Rather than filling in all the missing intermediate conclusions, we can see directly whether a sorites is valid by applying the rules we went over in Section 9. To spare you from flipping back and forth too much in the text, let's repeat the five basic rules here, modified so that they work with the longer arguments we are discussing.

- Rule** Each middle term in a valid categorical sorites argument must  
 1: be distributed at least once.  
**Rule** If a term is distributed in the conclusion of a valid categorical  
 2: sorites argument, then it must also be distributed in one of the  
 premises.  
**Rule** A valid Aristotelian syllogism cannot have two or more negative  
 3: premises.  
**Rule** A valid Aristotelian syllogism can have a negative conclusion if  
 4: and only if it has exactly one negative premise.  
**Rule** A valid Aristotelian syllogism with only universal premises  
 5: cannot have a particular conclusion.

Here "conclusion" means the final conclusion of the whole sorites, not any of the implicit intermediate conclusions. The middle terms are the

ones that appear in two premises. Also, as was the case with Aristotelian syllogisms, if Rule 5 is broken, the argument can be fixed by adding an appropriate existential premise. We will look at how to do that shortly.

To evaluate an sorites argument using the rules, it helps not only to have the argument in standard form, but also to mark each term as distributed or undistributed. Remember that a term is distributed in a statement if the statement is making a claim about that whole class. (See our original discussion on page 100.) Table 9.3 is a reminder of which kinds of sentence distribute which terms. Basically, universal statements distribute the subject, and negative statements distribute the predicate. In the example below, distributed terms are marked with a superscript D.

- (BF1) No  $B^D$  are  $A^D$ .  
 (BF2) All  $C^D$  are  $B$ .  
 (BF3) Some  $C$  are  $D$ .  
 (BF4) All  $E^D$  are  $D$ .  
 (BF5)  $\therefore$  All  $E^D$  are  $A$ .

<u>Mood</u>	<u>Form</u>	<u>Terms Distributed</u>
A	All $S$ are $P$	$S$
E	No $S$ are $P$	$S$ and $P$
I	Some $S$ are $P$	None
O	Some $S$ are not $P$	$P$

Table 9.3: Moods and distribution

To check the above argument using the rules we just run through the rules in order, and see if they are all followed. Rule 1 says that every middle term must be distributed in at least one premise. The middle terms here are  $B$ ,  $C$ , and  $D$ . We can quickly check to see that  $B$  is distributed in premise 1,  $C$  is distributed in premise 2, but  $D$  is never distributed. You can check this by trying to break the argument down into component syllogisms, as we did in the first section.

Sometimes the argument you are given will satisfy the two distribution rules but fail to satisfy other rules. Consider this example.

- (BG1) No  $C^D$  are  $B^D$ .  
 (BG2) Some  $A$  are  $B$ .  
 (BG3) No  $A^D$  are  $D^D$ .  
 (BG4)  $\therefore$  Some  $C$  are  $D$ .

Here the two middle terms are  $A$  and  $B$ , which are distributed in the third and first premises, respectively, so the first rule is satisfied. No terms are distributed in the conclusion, so neither need to be distributed in the premises, and the second rule is satisfied. This argument fails Rule 3, however, because it contains two negative premises. Again, you can confirm the results of this method using the results of the previous method.



The final case we need to look at is an argument that fails Rule 5, and thus needs an existential premise added.

(BH1) No  $A^D$  are  $D^D$ .

(BH2) All  $C^D$  are  $A$ .

(BH3) All  $C^D$  are  $B$ .

(BH4)  $\therefore$  Some  $B$  are not  $D^D$ .

Rules 1–4 are satisfied: the middle terms are  $A$  and  $C$ , which are both distributed;  $D$  is distributed in the conclusion and in  $P_1$ ; there is exactly one negative premise and the conclusion is also negative. However, Rule 5 is not satisfied, because all the premises are universal, but the conclusion is particular. This means that for the argument to be valid, we will need to add an existential premise.

But what existential premise should we add? Remember that the term that must exist in order for a conditionally valid argument to be actually valid is called the critical term (see page 124). When looking at conditionally valid arguments in Section 9, we basically just found the critical term by trial and error. For instance, when we used Venn diagrams to show that Felapton was conditionally valid (p. ??), we saw that adding either the premise “Some  $S$  exist” or the premise “Some  $P$  exist” would not help, but adding “Some  $M$  exist” would help. On the other hand, when considering the argument EII-3 (p. ??), we saw that there were no existential premises that we could add that would help. However, as the number of terms we have to deal with increases, the trial and error method becomes less appealing. On top of that, our previous trial and error method was guided by the Venn diagrams for the syllogisms, and we haven’t even looked yet at how to do Venn diagrams for sorites arguments.

The trick to finding the critical term in sorites arguments is called the superfluous distribution rule. The **SUPERFLUOUS DISTRIBUTION RULE** says that in a conditionally valid argument, the critical term will be the one that is distributed more times in the premises than is necessary to satisfy Rules 1 and 2. In the argument above, the middle term  $C$  is distributed in two premises, but it only needs to be distributed once to satisfy Rule 1. So we can make the argument above valid by adding the premise “Some  $C$  exist”:

(BI1) No  $A^D$  are  $D^D$ .

(BI2) All  $C^D$  are  $A$ .

(BI3) All  $C^D$  are  $B$ .

(BI4) Some  $C$  exist.

(BI5)  $\therefore$  Some  $B$  are not  $D^D$ .

You can confirm that this is a valid argument now using the method of filling in the missing intermediate conclusions described in the previous subsection. You will find that this argument actually consists of a Celarent

(EAE-1) and a Felapton (EAO-3). The latter is conditionally valid and has the critical term as its middle term.

### *A Venn Diagram with More Than Three Terms*

The last method we will consider is the most advanced but also the most fun. Your instructor may decide to skip this section, given that the prior two techniques are sufficient to evaluate sorites arguments. The Venn diagrams we have been making for this section have had three circles to represent three terms. The arguments we are dealing with have four or more terms, so if we want to represent them all at once, we will need a diagram with more shapes. To see how to set this up we need to remember the principles behind the original design of the diagram.

Recall that John Venn's original insight was that we should always begin by arranging the shapes so that every possible combination of the terms was represented (see pages ?? and ??). If we have two terms, we can use circles arranged like this,

[pages=-,pagecommand=,width=]figures/2venn0.pdf

where area 1 represents things that are  $S$  but not  $P$ , area 2 represents things that are both  $S$  and  $P$ , area 3 represents things that are only  $P$ , and area 4 represents things that are neither  $S$  nor  $P$ .

Three terms give us eight possibilities to represent, so we draw the diagram like this:

A four-term sorites will have 16 possible combinations of terms. To represent all of these, we will need to stretch out our circles into ellipses.

The diagram is complex, and it takes some practice to learn to work with it. However, the basic methods for using it are the same as with the three-term diagram. Consider this argument, taken from Lewis Carroll's logic textbook (?).

(Bj1) Nobody who is despised can manage a crocodile.

(Bj2) Illogical persons are despised.

(Bj3) Babies are illogical.

(Bj4)  $\therefore$  No babies can manage a crocodile.

If we set the major term "Persons who can manage a crocodile" as  $A$ , the minor term "babies" as  $D$ , and the two middle terms as  $B$  and  $C$ , we get this for the Venn diagram.

$A$ : Persons who can manage a crocodile

$B$ : Persons who are despised

*C*: Illogical persons

*D*: Babies



## **Part IV**

# **Propositional Logic**



## Propositional Logic

In Part III, we introduced a system of logic that dealt with categorical statements, statements like “All people are mortal” or “Some dogs have fleas.” The system developed there was somewhat formal, because it replaced some of the contents of ordinary English sentences with abstract symbols. In Part ??, we go the rest of the way, and replace all of ordinary English with abstract symbols, thus creating a fully artificial language. In the previous system, capital letters like  $S$  and  $P$  stood for categories, like “dogs” or “things that have fleas.” In the new system individual letters will stand for whole sentences, like “Tom wants to go to the bookstore” or “The sky is blue.” Because individual letters stand for sentences this kind of system is known as *sentential logic*, a term which we will be able to precisely define on page 177. We will call the specific version of sentential logic we will be developing SL.

### Sentence Letters

The most basic unit in our formal language SL is an individual capital letter— $A, B, C, D$ , etc. These letters, called **SENTENCE LETTERS**, are used to represent individual statements. Remember in section 1, we defined a statement as some bit of language that can be true or false, and listed all kinds of things that count as statements in English, from “*Tyrannosaurus rex* went extinct 65 million years ago” to “Lady Gaga is pretty.” In SL, all these statements are reduced to single capital letters.

Considered only as a symbol of SL, the letter  $A$  could mean any statement.

So when translating from English into SL, it is important to provide a **translation key**. Previously, we used translation keys to say assign the variables  $S$ ,  $M$ , and  $P$  to terms. (See page 159.) Now we will use them to assign sentences to sentence letters.

In order to specify what we mean, we need to provide a key saying what the sentence letters represent. We will call a list that assigns English phrases or sentences to variable names a **TRANSLATION KEY**. These are

sometimes also called “symbolization keys” or simply just “dictionaries.”

Consider this argument (recall that the portion of the passage in italics establishes the context, and is not part of the passage):

*A teacher is looking to see who has come to class* There is an apple on the desk.  
If there is an apple on the desk, then Jenny made it to class. Therefore, Jenny made it to class.

In canonical form, the argument would look like this:

- (A1) There is an apple on the desk.
- (A2) If there is an apple on the desk, then Jenny made it to class.
- (A3)  $\therefore$  Jenny made it to class.

A good symbolization key for this passage would look like this:

A: There is an apple on the desk.

B: Jenny made it to class.

Why do the symbolization key this way? The argument we are looking at is obviously valid in English. In symbolizing it, we want to preserve the structure of the argument that makes it valid. We could have made each sentence in the original argument into its own letter. Then the symbolization key would look like this:

A: There is an apple on the desk.

B: If there is an apple on the desk, then Jenny made it to class.

C: Jenny made it to class.

But that would mean the argument would look like this:

- (B1) A
- (B2) B
- (B3)  $\therefore$  C

There is no necessary connection between some sentence *A*, which could be any statement, and some other sentences *B* and *C*, which could also be anything. The structure of the argument has been completely lost in this translation.

The important thing about the argument is that the second premise is not merely *any* statement, logically divorced from the other statement in the argument. The second premise contains the first premise and the conclusion *as parts*. Our original symbolization key allows us to write the argument like this.

- (C1) A



(C2) If  $A$ , then  $B$ .

(C3)  $\therefore B$

This preserves the structure of the argument that makes it valid, but it still makes use of the English expression “If . . . then . . .” Although we ultimately want to replace all of the English expressions with logical notation, this is a good start.

The individual sentence letters in SL are called atomic sentences, because they are the basic building blocks out of which more complex sentences can be built. We can identify atomic sentences in English as well. An **ATOMIC SENTENCE** is one that cannot be broken into parts that are themselves sentences. “There is an apple on the desk” is an atomic sentence in English, because you can’t find any proper part of it that forms a complete sentence. For instance “an apple on the desk” is a noun phrase, not a complete sentence. Similarly “on the desk” is a prepositional phrase, and not a sentence, and “is an” is not any kind of phrase at all. This is what you will find no matter how you divide “There is an apple on the desk.” On the other hand you can find two proper parts of “If there is an apple on the desk, then Jenny made it to class” that are complete sentences: “There is an apple on the desk” and “Jenny made it to class.” As a general rule, we will want to use atomic sentences in SL (that is, the sentence letters) to represent atomic sentences in English. Otherwise, we will lose some of the logical structure of the English sentence, as we have just seen.

There are only 26 letters of the alphabet, but there is no logical limit to the number of atomic sentences. We can use the same letter to symbolize different atomic sentences by adding a subscript, a small number written after the letter. We could have a symbolization key that looks like this:

$A_1$ : The apple is under the armoire.

$A_2$ : Arguments in SL always contain atomic sentences.

$A_3$ : Adam Ant is taking an airplane from Anchorage to Albany.

:

$A_{294}$ : Alliteration angers otherwise affable astronauts.

Keep in mind that each of these is a different sentence letter. When there are subscripts in the symbolization key, it is important to keep track of them.

### *Sentential Connectives*

The previous section introduced the basic elements of SL, the sentence letters. But when we were looking at the argument involving Jenny and the apple, we saw that the best way to write a dictionary for the argument

left the words “if” and “then” in English. In this section we will introduce ways to connect the sentence letters together that will allow us to form a complete artificial language.

The symbols used to connect sentence letters are called **SENTENTIAL CONNECTIVES**, naturally enough. SL uses five sentential connectives:  $\wedge$ ,  $\vee$ ,  $\sim$ ,  $\rightarrow$ , and  $\leftrightarrow$ . To write the sentence about Jenny and the apple we use the symbol “ $\rightarrow$ .” Using the dictionary above, “If there is an apple on the desk, then Jenny made it to class” becomes  $A \rightarrow B$ . Table 10.1 summarizes the meaning of the five sentential connectives.

The sentential connectives are a kind of **LOGICAL CONSTANT**, because their meaning is fixed by the formal language that we have chosen. The other logical constants in SL are the parentheses. These are the things we cannot change in the symbolization key. The sentence letters, by contrast, are **NONLOGICAL SYMBOLS**, because their meaning can change as we change the symbolization key. We can decide that  $A$  stands for “Arthur is an aardvark” in one translation key and “Apu is an anthropologist” in the next. But we can’t say that the  $\sim$  symbol will mean “not” in one argument and “perhaps” in another.

The sections below describe each connective in more detail.

Symbol	What it is called	What it means
$\sim$	negation	“It is not the case that . . .”
$\wedge$	conjunction	“Both . . . and . . .”
$\vee$	disjunction	“Either . . . or . . .”
$\rightarrow$	conditional	“If . . . then . . .”
$\leftrightarrow$	biconditional	“. . . if and only if . . .”

Table 10.1: The Sentential Connectives.

## Negation

Consider how we might symbolize these sentences:

*not1* Mary is in Barcelona.

*not2* Mary is not in Barcelona.

*not3* Mary is somewhere other than Barcelona.

In order to symbolize sentence *not1*, we will need one sentence letter. We can provide a symbolization key:

$B$ : Mary is in Barcelona.

Note that here we are giving  $B$  a different interpretation than we did in the previous section. The symbolization key only specifies what  $B$  means *in a specific context*. It is vital that we continue to use this meaning of  $B$  so long as we are talking about Mary and Barcelona. Later, when we are

symbolizing different sentences, we can write a new symbolization key and use  $B$  to mean something else.

Now, sentence ?? is simply  $B$ . Sentence ?? is obviously related to sentence ??; it is basically ?? with a “not” added. We could put the sentence partly in our symbolic language by writing “Not  $B$ .” This means we do not want to introduce a different sentence letter for ?? . We just need a new symbol for the “not” part. Let’s use the symbol ‘ $\sim$ ,’ which we will call **NEGATION**. Now we can translate ‘Not  $B$ ’ to  $\sim B$ .

Sentence ?? is about whether or not Mary is in Barcelona, but it does not contain the word “not.” Nevertheless, it is obviously logically equivalent to sentence ?? . They both say that if you are looking for Mary, you shouldn’t look in Barcelona. Remember that in section 3, we said that two sentences in English are logically equivalent if they always have the same truth value. For our purposes, this means that they basically say the same thing. It is clear then that ?? and ?? are logically equivalent, so we can translate them both as  $\sim B$ .

Consider these further examples:

*not4* The widget can be replaced if it breaks.

*not5* The widget is irreplaceable.

*not5b* The widget is not irreplaceable.

If we let  $R$  mean “The widget is replaceable”, then sentence ?? can be translated as  $R$ . Sentence ?? means the opposite of sentence ??, so we can translate it  $\sim R$ . Sentence ?? adds another negation to sentence ?? . We know, as competent English speakers, that the two negations cancel each other out, so that sentence ?? is equivalent to sentence ?? . But the fact that two negations cancel each other out is a part of the logic of English that we actually want to capture with our formal language SL. So we will represent the two negations in sentence ?? as two negations in SL:  $\sim\sim R$ . We will now have to be sure that in SL the sentences  $R$  and  $\sim\sim R$  mean the same thing.

As the above examples begin to indicate, English has all kinds of ways to negate a sentence. Sometimes we use an explicit “not.” Sometimes we use a prefix like the “ir-” in “irreplaceable.” SL has just one way to form a negation: slap a  $\sim$  in front of the sentence. There is an English expression, however, that always occurs in the same place in an English sentence as the  $\sim$  occurs in the sentence SL. The English phrase is “It is not the case that.” Although this phrase sounds awkward, it always occurs in front of the sentence it is negating, just as the symbol  $\sim$  does. This makes it useful in translating sentences from SL back into English.  $\sim R$  can be translated “it is not the case that this widget is replaceable.” In the earlier example,  $\sim B$  can be translated “It is not the case that Mary is in Barcelona.”

Sometimes negations in English do not function as neatly as the  $\sim$  does in SL, because two things aren’t perfect opposites. Consider these

A sentence can be symbolized as  $\sim \mathcal{A}$  can always be paraphrased in English as “It is not the case that  $\mathcal{A}$ .”

sentences:

*not6* Elliott is happy.

*not7* Elliott is unhappy.

If we let  $H$  mean “Elliott is happy”, then we can symbolize sentence ?? as  $H$ , but does ?? really mean the same thing as  $\sim H$ ? Saying “Elliott is unhappy” indicates that Elliott is actively sad. But  $\sim H$  can be paraphrase as simply “It is not the case that Elliott is happy,” which might merely mean that Elliott is just feeling neutral. As we saw on page ??, the logics we discuss in this textbook are *bivalent*. Statements are only either true or false. Everything is in black and white, and issues like Elliott’s fine gradations in mood cannot be directly represented in our system. So in SL, sentences ?? and ?? would generally be represented by separate sentence letters.

One way of capturing the meaning of a sentential connective is to make a table which shows how the connective changes the meaning of the sentences it is applied to. The negation simply reverses the truth value of any sentence it is put in front of. For any sentence  $\mathcal{A}$ : If  $\mathcal{A}$  is true, then  $\sim \mathcal{A}$  is false. If  $\sim \mathcal{A}$  is true, then  $\mathcal{A}$  is false. Using T for true and F for false, we can summarize this in a *characteristic truth table* for negation:

$\mathcal{A}$	$\sim \mathcal{A}$
T	F
F	T

We will discuss truth tables at greater length in the next chapter.

## Conjunction

Consider these sentences:

*and1* Adam is athletic.

*and2* Barbara is athletic.

*and3* Adam is athletic, and Barbara is also athletic.

We will need separate sentence letters for ?? and ??, so we define this symbolization key:

$A$ : Adam is athletic.

$B$ : Barbara is athletic.

Sentence ?? can be symbolized as  $A$ . Sentence ?? can be symbolized as  $B$ . Sentence ?? can be paraphrased as “ $A$  and  $B$ .” In order to fully symbolize this sentence, we need another symbol. We will use  $\wedge$ . We translate “ $A$  and  $B$ ” as  $A \wedge B$ . The logical connective  $\wedge$  is called the **CONJUNCTION**, and  $A$  and  $B$  are each called **CONJUNCTS**.

Notice that we make no attempt to symbolize “also” in sentence ??.

Words like “both” and “also” function to draw our attention to the fact that two things are being conjoined. They are not doing any further logical work, so we do not need to represent them in SL.

Some more examples:

*and4* Barbara is athletic and energetic.

*and5* Barbara and Adam are both athletic.

*and6* Although Barbara is energetic, she is not athletic.

*and7* Barbara is athletic, but Adam is more athletic than she is.

Sentence ?? is obviously a conjunction. The sentence says two things about Barbara, so in English it is permissible to refer to Barbara only once. It might be tempting to try this when translating the argument: Since  $B$  means “Barbara is athletic”, one might paraphrase the sentences as “ $B$  and energetic.” This would be a mistake. Once we translate part of a sentence as  $B$ , any further structure is lost.  $B$  is an atomic sentence; it is nothing more than true or false. Conversely, “energetic” is not a sentence; on its own it is neither true nor false. We should instead paraphrase the sentence as “ $B$  and Barbara is energetic.” Now we need to add a sentence letter to the symbolization key. Let  $E$  mean “Barbara is energetic.” Now the sentence can be translated as  $B \wedge E$ .

Sentence ?? says one thing about two different subjects. It says of both Barbara and Adam that they are athletic, and in English we use the word “athletic” only once. In translating to SL, it is important to realize that the sentence can be paraphrased as, “Barbara is athletic, and Adam is athletic.” This translates as  $B \wedge A$ .

Sentence ?? is a bit more complicated. The word “although” sets up a contrast between the first part of the sentence and the second part. Nevertheless, the sentence says both that Barbara is energetic and that she is not athletic. In order to make each of the conjuncts an atomic sentence, we need to replace “she” with “Barbara.”

So we can paraphrase sentence ?? as, “*Both* Barbara is energetic, *and* Barbara is not athletic.” The second conjunct contains a negation, so we paraphrase further: “*Both* Barbara is energetic *and it is not the case that* Barbara is athletic.” This translates as  $E \wedge \sim B$ .

Sentence ?? contains a similar contrastive structure. It is irrelevant for the purpose of translating to SL, so we can paraphrase the sentence as “*Both* Barbara is athletic, *and* Adam is more athletic than Barbara.” (Notice that we once again replace the pronoun “she” with her name.) How should we translate the second conjunct? We already have the sentence letter  $A$  which is about Adam’s being athletic and  $B$  which is about Barbara’s being athletic, but neither is about one of them being more athletic than the other.

A sentence can be symbolized as  $\mathcal{A} \wedge \mathcal{B}$  if it can be paraphrased in English as ‘Both  $\mathcal{A}$ , and  $\mathcal{B}$ .’ Each of the conjuncts must be a sentence.

We need a new sentence letter. Let  $R$  mean “Adam is more athletic than Barbara.” Now the sentence translates as  $B \wedge R$ .

It is important to keep in mind that the sentence letters  $A$ ,  $B$ , and  $R$  are atomic sentences. Considered as symbols of SL, they have no meaning beyond being true or false. We have used them to symbolize different English language sentences that are all about people being athletic, but this similarity is completely lost when we translate to SL. No formal language can capture all the structure of the English language, but as long as this structure is not important to the argument there is nothing lost by leaving it out.

As with the negation, we can understand the meaning of the conjunction by making a table that shows how the conjunction affects the truth value of the sentences it is bringing together. For any sentences  $A$  and  $B$ ,  $A \wedge B$  is true if and only if both  $A$  and  $B$  are true. We can summarize this in the characteristic truth table for conjunction:

$A$	$B$	$A \wedge B$
T	T	T
T	F	F
F	T	F
F	F	F

Conjunction is symmetrical because we can swap the conjuncts without changing the truth value of the sentence. Regardless of what  $A$  and  $B$  are,  $A \wedge B$  is logically equivalent to  $B \wedge A$ .

## Disjunction

Consider these sentences:

or1 Either Denison will play golf with me, or he will watch movies.

or2 Either Denison or Ellery will play golf with me.

For these sentences we can use this symbolization key:

$D$ : Denison will play golf with me.

$E$ : Ellery will play golf with me.

$M$ : Denison will watch movies.

Sentence ?? is “Either  $D$  or  $M$ .” To fully symbolize this, we introduce a new symbol. The sentence becomes  $D \vee M$ . The  $\vee$  connective is called **DISJUNCTION**, and  $D$  and  $M$  are called **DISJUNCTS**.

Sentence ?? is only slightly more complicated. There are two subjects, but the English sentence only gives the verb once. In translating, we can

Sentences that can be paraphrased “ $A$ , but  $B$ ” or “*Although*  $A$ ,  $B$ ” are best symbolized using conjunction  $A \wedge B$ .

paraphrase it as “Either Denison will play golf with me, or Ellery will play golf with me.” Now it obviously translates as  $D \vee E$ .

The English word “or” is somewhat ambiguous. Sometimes in English, when we say “this or that,” we mean that either option is possible, but not both. For instance, if a restaurant menu says, “Entrées come with either soup or salad” we naturally assume you can have soup, or you can have salad; but, if you want *both* soup *and* salad, then you will have to pay extra. This kind of disjunction is called an **EXCLUSIVE OR**, because it excludes the possibility that both disjuncts are true.

At other times, the word “or” allows for the possibility that both disjuncts might be true. This is probably the case with sentence ??, above. I might play with Denison, with Ellery, or with both Denison and Ellery. Sentence ?? merely says that I will play with *at least* one of them. The **INCLUSIVE OR** is the kind of disjunction that allows for the possibility that both disjuncts are true. The inclusive or says “This or that, or both.”

To goal of a formal language is to remove ambiguity, so we need to pick one of these ors. SL follows tradition and uses the symbol  $\vee$  to represent an *inclusive or*. This winds up being reflected in the characteristic truth table for the  $\vee$ . The sentence  $D \vee E$  is true if  $D$  is true, if  $E$  is true, or if both  $D$  and  $E$  are true. It is false only if both  $D$  and  $E$  are false. The truth table looks like this:

A	B	$A \vee B$
T	T	T
T	F	T
F	T	T
F	F	F

Like conjunction, disjunction is symmetrical.

$$\mathcal{A} \vee$$

*is logically equivalent to*  $\mathcal{B} \vee \mathcal{A}$ .

## Conditional

We already met the conditional at the start of this section, when we were discussing the sentence “If there is an apple on the table, Jenny made it to class,” which became  $A \rightarrow B$ . The symbol  $\rightarrow$  is called a **CONDITIONAL**. The sentence on the left-hand side of the conditional ( $A$  in this example) is called the **ANTECEDENT**. The sentence on the right-hand side ( $B$ ) is called the **CONSEQUENT**.

Like the English word “or,” the English phrase “if...then...” has some ambiguity. Consider our original example, “If there is an apple on the table, Jenny made it to class.” The statements tells us what we should infer if there is an apple on the table, but what if there *isn't* an apple on the table. Does

A sentence can be symbolized as  $\mathcal{A} \vee \mathcal{B}$  if it can be paraphrased in English as “Either  $\mathcal{A}$  or  $\mathcal{B}$ .” Each of the disjuncts must be a sentence.

that guarantee that Jenny did not make it to class? It could be that an apple on the table is a clear sign that Jenny made it to class, because no one else would put an apple on the table, but nevertheless Jenny sometimes comes to class without putting an apple on the table.

We can get a good sense of the decision we face if we try to write up the characteristic truth table for the conditional. The first two lines are easy. The sentence “If  $\mathcal{A}$ , then  $\mathcal{B}$ ” means that if  $\mathcal{A}$  is true, then  $\mathcal{B}$  is true. This would be confirmed by the situation where both  $\mathcal{A}$  and  $\mathcal{B}$  are true, but falsified by the situation where  $\mathcal{A}$  is true and  $\mathcal{B}$  is false. In terms of our example, if we came to class and found the apple there, but Jenny absent, we would know that the statement “If there is an apple on the table, Jenny made it to class” is false. But if we came to class and found both Jenny and the apple present, we could say that the statement “If there is an apple on the table, Jenny made it to class” is true. That gives us this much of a truth table.

$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A} \rightarrow \mathcal{B}$
T	T	T
T	F	F
F	T	?
F	F	?

How do we fill in the question marks in the last two lines? In real life, we would generally make judgments on a case by case basis, relying heavily on the context we are in. But for a formal language we just want to lay down a simple rule. The traditional solution for sentential logic is to say that the conditional is what logicians call a “material conditional.” If the antecedent of a material conditional is false, then the whole statement is automatically true, regardless of the truth value of  $\mathcal{B}$ . In short,  $\mathcal{A} \rightarrow \mathcal{B}$  is false if and only if  $\mathcal{A}$  is true and  $\mathcal{B}$  is false. We can summarize this with a characteristic truth table for the conditional.

$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A} \rightarrow \mathcal{B}$
T	T	T
T	F	F
F	T	T
F	F	T

The conditional is asymmetrical. You cannot swap the antecedent and consequent without changing the meaning of the sentence, because  $\mathcal{A} \rightarrow \mathcal{B}$  and  $\mathcal{B} \rightarrow \mathcal{A}$  are not logically equivalent.

Not all sentences of the form “If . . . , then . . .” are conditionals. Consider this sentence:

*if5* If anyone wants to see me, then I will be on the porch.

When I say this, it means that I will be on the porch, regardless of whether anyone wants to see me or not—but if someone did want to see me,



then they should look for me there. If we let  $P$  mean “I will be on the porch,” then sentence ?? can be translated simply as  $P$ .

### *Biconditional*

The conditional was an asymmetric connective. The sentence  $A \rightarrow B$  does not mean the same thing as the sentence  $B \rightarrow A$ . It is convenient to have a single symbol that combines the meaning of these two sentences. The **BICONDITIONAL**—written as double headed arrow,  $\leftrightarrow$ —is a sentential connective used to represent a situation where  $A$  implies  $B$  and  $B$  implies  $A$ .

To draw up the characteristic truth table for the biconditional, we need to think about the situations where  $A \rightarrow B$  and  $B \rightarrow A$  are false. The sentence  $A \rightarrow B$  is only false when  $A$  is true and  $B$  is false. For  $B \rightarrow A$  the reverse is true. It is false when  $B$  is true and  $A$  is false. Our biconditional  $A \leftrightarrow B$  needs to avoid both of these situations to be true, because it is only true when  $A \rightarrow B$  and  $B \rightarrow A$  are true. This, then, is the characteristic truth table for the biconditional. It says that the biconditional is true when the truth values of the two sides match.

$A$	$B$	$A \leftrightarrow B$
T	T	T
T	F	F
F	T	F
F	F	T

If the biconditional holds between two sentences, we can say that the two sentences are logically equivalent. Back on page 60, we said that two sentences were logically equivalent if they always have the same truth value. That is exactly what is happening here.

Back in section 7, we saw that the system of categorical logic we were studying at the time could actually represent a large range of sentences in ordinary English, even though it only had the quantifiers “All” and “some” plus negation. In this section, we will see that something similar happens with SL. There is actually a lot we can cover, even though we only have five connectives.

The previous section introduced the five sentential connectives. Now we will look at some trickier translations involving those connectives.

### *Combining connectives*

In our system of categorical logic, we just had four kinds of sentences—A, E, I, and O—and if we wanted to combine them, the only way to do that would be to form a syllogism. In SL, we can combine an unlimited number of connectives together into a single sentence to express complicated ideas

that couldn't be represented by Aristotelean logic. A single sentence in SL can use multiple connectives.

Consider the English sentence "If it is not raining, we will have a picnic." There are two aspects of this sentence we will want to represent with sentential connectives in SL, the "if...then..." structure and the negation in the first part of the sentence. The rest of the sentence can be represented by these sentence letters

*A*: It is raining.

*M*: We will have a picnic.

We can then translate the whole sentence into SL like this:  $\sim A \rightarrow B$ . We can make sentences as complicated as we want this way, even to the point where the equivalent English sentence would be impossible to follow. The sentence  $\sim(P \wedge Q) \rightarrow [(R \vee S) \leftrightarrow \sim(T \wedge U)]$  is perfectly acceptable in SL, even if any English sentence it translates into would be a monster. This is part of the power of a complete formal language like SL, but it is also why arguments in SL begin to resemble the ob/ob mouse more than they resemble any argument you might encounter in the wild. (See page 26)

Although sentences in SL can be as long as you like, you can't just combine symbols any old way. There is a specific set of rules you have to follow. These are outlined in section 10, below.

The fact that we can write these more complicated sentences means we can actually do without some of the connectives we have given ourselves in SL. For instance, we don't really need the biconditional. Any sentence of the form  $\mathcal{A} \leftrightarrow \mathcal{B}$  is going to be equivalent to the sentence  $(\mathcal{A} \rightarrow \mathcal{B}) \wedge (\mathcal{B} \rightarrow \mathcal{A})$ . This just follows from the way we defined the biconditional earlier. Nevertheless, tradition and convenience mandate that we give the biconditional a separate symbol.

## Unless

Because our connectives can be put together in different ways, some English sentences can be represented equally well by multiple sentences in SL. English sentences involving the word "unless" are a case in point.

*unless1* Unless you wear a jacket, you will catch cold.

*unless2* You will catch cold unless you wear a jacket.

These are basically two different version of the same English sentence. The only difference is that in one case, the "unless" clause comes first, and in the other it comes second. Let *J* mean "You will wear a jacket" and let *C* mean "You will catch a cold." We can paraphrase sentence ?? as "Unless *J*, *C*." This means that if you do not wear a jacket, then you will catch cold. With this in mind, we might translate it as  $\sim J \rightarrow C$ . It also means that

if you do not catch a cold, then you must have worn a jacket; with this in mind, we might translate it as  $\sim C \rightarrow J$ .

Which of these is the correct translation of sentence ??? Both translations are correct, because the two translations are logically equivalent in SL. Sentence ??, in English, is logically equivalent to sentence ??. So, it also can be translated as either  $\sim J \rightarrow D$  or  $\sim D \rightarrow J$ .

When symbolizing sentences like sentence ?? and sentence ??, it is easy to get turned around. We have two different versions of the English sentence and two different versions of the sentence in SL. The important thing to see here is that none of these sentences are equivalent to  $J \rightarrow \sim D$ . *The negated statement must be the antecedent to the conditional.*

If this is too many options to keep track of, there is a simpler alternative. It turns out that any “unless” statement is actually equivalent to an “or” statement. Both statements ?? and ?? mean that you will wear a jacket or—if you do not wear a jacket—then you will catch a cold. So we can translate them as  $J \vee D$ . (You might worry that the “or” here should be an *exclusive or*. However, the sentences do not exclude the possibility that you might *both* wear a jacket *and* catch a cold; jackets do not protect you from all the possible ways that you might catch a cold.)

If a sentence can be paraphrased as “Unless  $\mathcal{A}, \mathcal{B}$ ,” then it can be symbolized as  $\mathcal{A} \vee \mathcal{B}$ .

## Only

In section 7, we saw that the word “only” could reverse the meaning of a statement in Mood A. “All dogs are mammals” means something different than “Only dogs are mammals,” the first one is true but the second one is false. Something similar happens with conditional statements in SL. The word “only” can reverse the meaning of a conditional sentence in SL. For the following sentences, let  $R$  mean “You will cut the red wire” and  $B$  mean “The bomb will explode.”

*if1* If you cut the red wire, then the bomb will explode.

*if2* The bomb will explode only if you cut the red wire.

Sentence ?? can be translated partially as “If  $R$ , then  $B$ .” Sentence ?? is also a conditional. Since the word “if” appears in the second half of the sentence, it might be tempting to symbolize this in the same way as sentence ??. That would be a mistake.

The conditional  $R \rightarrow B$  says that *if*  $R$  were true, *then*  $B$  would also be true. It does not say that you cutting the red wire is the *only* way that the bomb could explode. Someone else might cut the wire, or the bomb might be on a timer. The sentence  $R \rightarrow B$  does not say anything about what to expect if  $R$  is false. Sentence ?? is different. It says that the only conditions under which the bomb will explode involve you having cut the red wire; i.e., if the bomb explodes, then you must have cut the wire. As such, sentence ?? should be symbolized as  $B \rightarrow R$ .

It is important to remember that the connective  $\rightarrow$  says only that, if the antecedent is true, then the consequent is true. It says nothing about the *causal* connection between the two events. Translating sentence ?? as  $B \rightarrow R$  does not mean that the bomb exploding would somehow have caused you cutting the wire. Both sentence ?? and ?? suggest that, if you cut the red wire, you cutting the red wire would be the cause of the bomb exploding. They differ on the *logical* connection. If sentence ?? were true, then an explosion would tell us—those of us safely away from the bomb—that you had cut the red wire. Without an explosion, sentence ?? tells us nothing.

The paraphrased sentence “*AonlyifB*” is logically equivalent to “If *A*, then *B*.”

Things can get a bit more complicated, because English also allows you to reverse the order of the clauses. Think about this sentence

*iff3* The bomb will explode, if you cut the red wire

This is just sentence ?? with the order of the clauses reversed, so it still means  $R \rightarrow B$ . Changing the order of the English clauses does not change the sentence in SL, but adding the word “only” does.

If this gets confusing, just remember this rule:

“If...” introduces the antecedent. “Only if...” introduces the consequent.

Because “if” and “only if” have opposite meanings, when we put them together, we get the biconditional. Consider these sentences:

*iff1* The figure on the board is a triangle only if it has exactly three sides.

*iff2* The figure on the board is a triangle if it has exactly three sides.

*iff3* The figure on the board is a triangle if and only if it has exactly three sides.

Let *T* mean “The figure is a triangle” and *S* mean “The figure has three sides.” Sentence ??, for reasons discussed above, can be translated as  $T \rightarrow S$ . Sentence ?? is importantly different. It can be paraphrased as “If the figure has three sides, then it is a triangle.” So it can be translated as  $S \rightarrow T$ .

Sentence ?? says that *T* is true *if and only if* *S* is true; we can infer *S* from *T*, and we can infer *T* from *S*. In other words, ?? is equivalent to  $T \rightarrow S$  and  $S \rightarrow T$ , which is the same as  $T \leftrightarrow S$

A final way to think about the way “only” effects a conditional sentence is to think about the difference between necessary and sufficient conditions. In a way, the terms are pretty much self explanatory. Nevertheless, it is really easy to get them confused, to the extent that even professional logicians and trained philosophers can get them mixed up.

A **NECESSARY CONDITION** is one that is needed for something else to be true, just like the name says. Having gas in the tank is a *necessary* condition for the car to move. It just doesn’t go anywhere without gas. However, having gas in the tank isn’t *all you need* to get the car moving. You also have to put the key in the ignition and turn it.

A **SUFFICIENT CONDITION**, on the other hand, is *all you need* for something else to be true. If something is a dog, that is a *sufficient* condition for it to be a mammal. Once you know Cupcake (Fig. 10.1) is a dog, you have enough information to infer that she is a mammal. Being a dog is not a necessary condition for being a mammal however. You can also be a mammal being being a cat, or a human, or a wombat.



Figure 10.1: This is Cupcake. The fact that she is a dog is a *sufficient* condition for her to be a mammal. She also likes socks.

The conditional symbol in SL represents a sufficient condition, at least when read forward. That is, the antecedent is a sufficient condition for the consequent. If you have the antecedent, that is all you need to know to infer the consequent. So if  $D$  is “Cupcake is a dog” and  $M$  is “Cupcake is a mammal”, then  $D \rightarrow C$  is true. Being a dog is sufficient for being a mammal. As it turns out, if the relationship is sufficient going one direction, it is necessary going the other. So being a mammal is a necessary condition for being a dog. If cupcake weren’t a mammal, there would be no way for her to be a dog. Figure 10.2 shows this relationship.

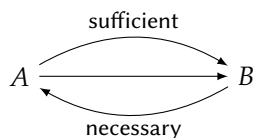


Figure 10.2: The antecedent of a material conditional is a sufficient condition for the consequent, while the consequent is a necessary condition for the antecedent.

### *Combining negation with conjunction and disjunction*

Tricky things happen when you combine a negation with a conjunction or disjunction, so it is worth taking a closer look here. Consider these sentences

*or3* Either you will not have soup, or you will not have salad.

*or4* You will have neither soup nor salad.

We let  $S_1$  mean that you get soup and  $S_2$  mean that you get salad. Sentence ?? can be paraphrased in this way: “Either *it is not the case that* you get soup, or *it is not the case that* you get salad.” Translating this requires both disjunction and negation. It becomes  $\sim S_1 \vee \sim S_2$ .

Sentence ?? also requires negation. It can be paraphrased as, “*It is not the case that* either you get soup or you get salad.” We need some way of indicating that the negation does not just negate the right or left disjunct, but rather negates the entire disjunction. In order to do this, we put parentheses around the disjunction: “It is not the case that  $(S_1 \vee S_2)$ .” This becomes simply  $\sim(S_1 \vee S_2)$ . Notice that the parentheses are doing important work here. The sentence  $\sim S_1 \vee S_2$  would mean “Either you will not have soup, or you will have salad.”

Something similar happens with negation and conjunction. Consider these sentences

*notand1* You can’t have soup and you can’t have salad.

*notand2* You can’t have both soup and salad.

In sentence ??, the two parts of the sentence are negated individually. We would translate it into SL like this:  $\sim S_1 \wedge \sim S_2$ . In sentence ??, the negation applies to soup and salad taken together. You are allowed to have soup only, or salad only. You just can’t have both together. We would translate sentence ?? like this:  $\sim(S_1 \wedge S_2)$ .

You can combine disjunction, conjunction, and negation to represent the exclusive or, as in this sentence.

*or.xor* You get either soup or salad, but not both.

Remember on page ??, we said that the  $\vee$  in SL represented an inclusive or. It said “this or that or both.” If we want to represent an exclusive or, we need to combine disjunction, conjunction and negation. We can break the sentence into two parts. The first part says that you get one or the other. We translate this as  $(S_1 \vee S_2)$ . The second part says that you do not get both. We can paraphrase this as “It is not the case both that you get soup and that you get salad.” Using both negation and conjunction, we translate this as  $\sim(S_1 \wedge S_2)$ . Now we just need to put the two parts together. As we saw above, “but” can usually be translated as a conjunction. Sentence ?? can thus be translated as  $(S_1 \vee S_2) \wedge \sim(S_1 \wedge S_2)$ .

## *Recursive Syntax for SL*

The previous two sections gave you a rough, informal sense of how to create sentences in SL. If I give you an English sentence like “Grass is either green or brown,” you should be able to write a corresponding sentence in SL: “ $A \vee B$ .” In this section we want to give a more precise definition of a sentence in SL. When we defined sentences in English, we did so using the concept of truth: Sentences were units of language that can be true or false. (See page 27.) In SL, it is possible to define what counts as a sentence without talking about truth. Instead, we can just talk about the structure of the sentence. This is one respect in which a formal language like SL is more precise than a natural language like English.

The structure of a sentence in SL considered without reference to truth or falsity is called its syntax. More generally **SYNTAX** refers to the study of the properties of language that are there even when you don’t consider meaning. Whether a sentence is true or false is considered part of its meaning. In this chapter, we will be giving a purely syntactical definition of a sentence in SL. The contrasting term is **SEMANTICS** the study of aspects of language that relate to meaning, including truth and falsity. (The word “semantics” comes from the Greek word for “mark”)

If we are going to define a sentence in SL just using syntax, we will need to carefully distinguish SL from the language that we use to talk about SL. When you create an artificial language like SL, the language that you are creating is called the **OBJECT LANGUAGE**. The language that we use to talk about the object language is called the **METALANGUAGE**. Imagine building a house. The object language is like the house itself. It is the thing we are building. While you are building a house, you might put up scaffolding around it. The scaffolding isn’t part of the house. You just use it to build the house. The metalanguage is like the scaffolding.

The object language in this chapter is SL. For the most part, we can build this language just by talking about it in ordinary English. However we will also have to build some special scaffolding that is not a part of SL, but will help us build SL. Our metalanguage will thus be ordinary English plus this scaffolding.

An important part of the scaffolding are the **METAVARIABLES**. These are the fancy script letters we have been using in the characteristic truth tables for the connectives:  $\mathcal{A}$ ,  $\mathcal{B}$ ,  $\mathcal{C}$ , etc. These are letters that can refer to any sentence in SL. They can represent sentences like  $P$  or  $Q$ , or they can represent longer sentences, like  $((\mathcal{A} \vee \mathcal{B}) \wedge \mathcal{G}) \rightarrow (\mathcal{P} \leftrightarrow \mathcal{Q})$ . Just as the sentence letters  $A$ ,  $B$ , etc. are variables that range over any English sentence, the metavariables  $\mathcal{A}$ ,  $\mathcal{B}$ , etc. are variables that range over any sentence in SL, including the sentence letters  $A$ ,  $B$ , etc.

As we said, in this chapter we will give a syntactic definition for “sentence of SL.” The definition itself will be given in mathematical English, the

metalanguage. Table 10.2 gives the basic elements of SL.

Element	Symbols
sentence letters	$A, B, C, \dots, Z, A_1, B_1, Z_1, A_2, A_{25}, J_{375}, \dots$
connectives	$\sim, \wedge, \vee, \rightarrow, \leftrightarrow$
parentheses	$(, )$

Table 10.2: The basic elements of SL

Most random combinations of these symbols will not count as sentences in SL. Any random connection of these symbols will just be called a “string” or “expression.” Random strings only become meaningful sentences when they are structured according to the rules of syntax. We saw from the earlier two sections that individual sentence letters, like  $A$  and  $G_{13}$  counted as sentences. We also saw that we can put these sentences together using connectives so that  $\sim A$  and  $\sim G_{13}$  is a sentence. The problem is, we can’t simply list all the different sentences we can put together this way, because there are infinitely many of them. Instead, we will define a sentence in SL by specifying the process by which they are constructed.

Consider negation: Given any sentence  $\mathcal{A}$  of SL,  $\sim \mathcal{A}$  is a sentence of SL. It is important here that  $\mathcal{A}$  is not the sentence letter  $A$ . Rather, it is a metavariable: part of the metalanguage, not the object language. Since  $\mathcal{A}$  is not a symbol of SL,  $\sim \mathcal{A}$  is not an expression of SL. Instead, it is an expression of the metalanguage that allows us to talk about infinitely many expressions of SL: all of the expressions that start with the negation symbol.

We can say similar things for each of the other connectives. For instance, if  $\mathcal{A}$  and  $\mathcal{B}$  are sentences of SL, then  $(\mathcal{A} \wedge \mathcal{B})$  is a sentence of SL. Providing clauses like this for all of the connectives, we arrive at the following formal definition for a **SENTENCE OF SL**:

1. Every atomic sentence is a sentence.
2. If  $\mathcal{A}$  is a sentence, then  $\sim \mathcal{A}$  is a sentence of SL.
3. If  $\mathcal{A}$  and  $\mathcal{B}$  are sentences, then  $(\mathcal{A} \wedge \mathcal{B})$  is a sentence.
4. If  $\mathcal{A}$  and  $\mathcal{B}$  are sentences, then  $(\mathcal{A} \vee \mathcal{B})$  is a sentence.
5. If  $\mathcal{A}$  and  $\mathcal{B}$  are sentences, then  $(\mathcal{A} \rightarrow \mathcal{B})$  is a sentence.
6. If  $\mathcal{A}$  and  $\mathcal{B}$  are sentences, then  $(\mathcal{A} \leftrightarrow \mathcal{B})$  is a sentence.
7. All and only sentences of SL can be generated by applications of these rules.

We can apply this definition to see whether an arbitrary string is a sentence. Suppose we want to know whether or not  $\sim \sim \sim D$  is a sentence of SL. Looking at the second clause of the definition, we know that  $\sim \sim \sim D$  is a sentence if  $\sim \sim D$  is a sentence. So now we need to ask whether or not  $\sim \sim D$  is a sentence. Again looking at the second clause of the definition,  $\sim \sim D$



is a sentence if  $\sim D$  is. Again,  $\sim D$  is a sentence if  $D$  is a sentence. Now  $D$  is a sentence letter, an atomic sentence of SL, so we know that  $D$  is a sentence by the first clause of the definition. So for a compound formula like  $\sim \sim \sim D$ , we must apply the definition repeatedly. Eventually we arrive at the atomic sentences from which the sentence is built up.

Definitions like this are called recursive. **RECURSIVE DEFINITIONS** begin with some specifiable base elements and define ways to indefinitely compound the base elements. Just as the recursive definition allows complex sentences to be built up from simple parts, you can use it to decompose sentences into their simpler parts. To determine whether or not something meets the definition, you may have to refer back to the definition many times. Recursive definitions are also sometimes called “inductive definitions.”

We are now in a position to define what it means for a system of logic to be a system of sentential logic. A **SENTENTIAL LOGIC** is a system of logic in which statements can be defined using a recursive definition with only sentences in the base class. This book defines one system of sentential logic, which we call SL. Other books use other systems.

When you use a connective to build a longer sentence from shorter ones, the shorter sentences are said to be in the **SCOPE** of the connective. So in the sentence  $(A \wedge B) \rightarrow C$ , the scope of the connective  $\rightarrow$  includes  $(A \wedge B)$  and  $C$ . In the sentence  $\sim(A \wedge B)$  the scope of the  $\sim$  is  $(A \wedge B)$ . On the other hand, in the sentence  $\sim A \wedge B$  the scope of the  $\sim$  is just  $A$ .

The last connective that you add when you assemble a sentence using the recursive definition is the **MAIN CONNECTIVE** of that sentence. For example: The main logical operator of  $\sim(E \vee (F \rightarrow G))$  is negation,  $\sim$ . The main logical operator of  $(\sim E \vee (F \rightarrow G))$  is disjunction,  $\vee$ . The main connective of any sentence will have all the rest of the sentence in its scope.

Because statement in our language is defined recursively, we can say it is “uniquely readable.” **UNIQUE READABILITY** is a property of formal languages

which is present when each part:quantifier, logic well formed formula, statement can only be constructed in a single way. Every process of building

### *Notational conventions*

A sentence like  $(Q \wedge R)$  must be surrounded by parentheses, because we might apply the definition again to use this as part of a more complicated sentence. If we negate  $(Q \wedge R)$ , we get  $\sim(Q \wedge R)$ . If we just had  $Q \wedge R$  without the parentheses and put a negation in front of it, we would have  $\sim Q \wedge R$ . It is most natural to read this as meaning the same thing as  $(\sim Q \wedge R)$ , something very different than  $\sim(Q \wedge R)$ . The sentence  $\sim(Q \wedge R)$  means that it is not the case that both  $Q$  and  $R$  are true;  $Q$  might be false or  $R$  might be false, but the sentence does not tell us which. The sentence  $(\sim Q \wedge R)$  means specifically that  $Q$  is false and that  $R$  is true. As such, parentheses are crucial to the meaning of the sentence.

So, strictly speaking,  $Q \wedge R$  without parentheses is *not* a sentence of SL. When using SL, however, we will often be able to relax the precise definition so as to make things easier for ourselves. We will do this in several ways.

First, we understand that  $Q \wedge R$  means the same thing as  $(Q \wedge R)$ . As a matter of convention, we can leave off parentheses that occur *around the entire sentence*.

Second, it can sometimes be confusing to look at long sentences with many nested pairs of parentheses. We adopt the convention of using square brackets [ and ] in place of parentheses. There is no logical difference between  $(P \vee Q)$  and  $[P \vee Q]$ , for example. The unwieldy sentence

$$(((H \rightarrow I) \vee (I \rightarrow H)) \wedge (J \vee K))$$

could be written in this way:

$$[(H \rightarrow I) \vee (I \rightarrow H)] \wedge (J \vee K)$$

Third, we will sometimes want to translate the conjunction of three or more sentences. For the sentence “Alice, Bob, and Candice all went to the party,” suppose we let  $A$  mean “Alice went,”  $B$  mean “Bob went,” and  $C$  mean “Candice went.” The definition only allows us to form a conjunction out of two sentences, so we can translate it as  $(A \wedge B) \wedge C$  or as  $A \wedge (B \wedge C)$ . There is no reason to distinguish between these, since the two translations are logically equivalent. There is no logical difference between the first, in which  $(A \wedge B)$  is conjoined with  $C$ , and the second, in which  $A$  is conjoined with  $(B \wedge C)$ . So we might as well just write  $A \wedge B \wedge C$ . As a matter of convention, we can leave out parentheses when we conjoin three or more sentences.

Fourth, a similar situation arises with multiple disjunctions. “Either Alice, Bob, or Candice went to the party” can be translated as  $(A \vee B) \vee C$  or as  $A \vee (B \vee C)$ . Since these two translations are logically equivalent, we may write  $A \vee B \vee C$ .

These latter two conventions only apply to multiple conjunctions or multiple disjunctions. If a series of connectives includes both disjunctions and conjunctions, then the parentheses are essential; as with  $(A \wedge B) \vee C$  and  $A \wedge (B \vee C)$ . The parentheses are also required if there is a series of conditionals or biconditionals; as with  $(A \rightarrow B) \rightarrow C$  and  $A \leftrightarrow (B \leftrightarrow C)$ .

We have adopted these four rules as notational conventions, not as changes to the definition of a sentence. Strictly speaking,  $A \vee B \vee C$  is still not a sentence. Instead, it is a kind of shorthand. We write it for the sake of convenience, but we really mean the sentence  $(A \vee (B \vee C))$ .

If we had given a different definition for a sentence, then these could count as sentences. We might have written rule 3 in this way: “If  $\mathcal{A}, \mathcal{B}, \dots \mathcal{Z}$  are sentences, then  $(\mathcal{A} \wedge \mathcal{B} \wedge \dots \wedge \mathcal{Z})$ , is a sentence.” This would make it easier to translate some English sentences, but would have the cost of

making our formal language more complicated. We would have to keep the complex definition in mind when we develop truth tables and a proof system. We want a logical language that is expressively simple and allows us to translate easily from English, but we also want a formally simple language. Adopting notational conventions is a compromise between these two desires.

### *Key Terms*

- |                             |                                  |
|-----------------------------|----------------------------------|
| 1. <b>Antecedent</b>        | 16. <b>Necessary condition</b>   |
| 2. <b>Atomic sentence</b>   | 17. <b>Negation</b>              |
| 3. <b>Biconditional</b>     | 18. <b>Nonlogical symbol</b>     |
| 4. <b>Conditional</b>       | 19. <b>Object language</b>       |
| 5. <b>Conjunct</b>          | 20. <b>Recursive definition</b>  |
| 6. <b>Conjunction</b>       | 21. <b>Scope</b>                 |
| 7. <b>Consequent</b>        | 22. <b>Semantics</b>             |
| 8. <b>Disjunct</b>          | 23. <b>Sentence letter</b>       |
| 9. <b>Disjunction</b>       | 24. <b>Sentence of SL</b>        |
| 10. <b>Exclusive or</b>     | 25. <b>Sentential connective</b> |
| 11. <b>Inclusive or</b>     | 26. <b>Sufficient condition</b>  |
| 12. <b>Logical constant</b> | 27. <b>Symbolization key</b>     |
| 13. <b>Main connective</b>  | 28. <b>Syntax</b>                |
| 14. <b>Metalanguage</b>     | 29. <b>Translation key</b>       |
| 15. <b>Metavariables</b>    |                                  |



## Truth Tables

This chapter introduces a way of evaluating sentences and arguments of SL called the truth table method. As we shall see, the truth table method is *semantic* because it involves one aspect of the meaning of sentences, whether those sentences are true or false. As we saw on page 175, semantics is the study of aspects of language related to meaning, including truth and falsity. Although it can be laborious, the truth table method is a purely mechanical procedure that requires no intuition or special insight. When we get to Chapter ??, we will provide a parallel semantic method for QL; however, this method will not be purely mechanical.

### Basic Concepts

In the previous chapter, we said that a formal language is built from two kinds of elements: logical constants and nonlogical symbols. The **LOGICAL CONSTANTS** have their meaning fixed by the formal language, while the **NONLOGICAL SYMBOLS** get their meaning in the symbolization key. The logical constants in SL are the sentential connectives and the parentheses, while the nonlogical symbols are the sentence letters.

When we assign meaning to the nonlogical symbols of a language using a dictionary, we say we are giving an “interpretation” of the language. More formally an **INTERPRETATION** of a language is a correspondence between elements of the object language and elements of some other language or logical structure. The symbolization keys we defined in Chapter ?? (p. 159) are one sort of interpretation. Fancier languages will have more complicated kinds of interpretations.

The truth table method will also involve giving an interpretation of sentences, but they will be much simpler than the translation keys we used in Chapter ?. We will not be concerned with what the individual sentence letters mean. We will only care whether they are true or false. In other words, our interpretations will assign **truth values** to the sentence letters. (See page 58.)

We can get away with only worrying about the truth values of sentence

letters because of the way that the meaning of larger sentences is generated by the meaning of their parts. Any larger sentence of SL is composed of atomic sentences with sentential connectives. The truth value of the compound sentence depends only on the truth value of the atomic sentences that it comprises. In order to know the truth value of  $D \leftrightarrow E$ , for instance, you only need to know the truth value of  $D$  and the truth value of  $E$ . Connectives that work in this way are called truth functional. More technically, we define a **TRUTH-FUNCTIONAL CONNECTIVE** as an operator that builds larger sentences out of smaller ones, and fixes the truth value of the resulting sentence based only on the truth value of the component sentences.

Because all of the logical symbols in SL are truth functional, we can study the semantics of SL looking only at truth and falsity. If we want to know about the truth of the sentence  $A \wedge B$ , the only thing we need to know is whether  $A$  and  $B$  are true. It doesn't actually matter what else they mean. So if  $A$  is false, then  $A \wedge B$  is false no matter what false sentence  $A$  is used to represent. It could be "I am the Pope" or "Pi is equal to 3.19." The larger sentence  $A \wedge B$  is still false. So to give an interpretation of sentences in SL, all we need to do is create a truth assignment. A **TRUTH ASSIGNMENT** is a function that maps the sentence letters in SL onto our two truth values. In other words, we just need to assign Ts and Fs to all our sentence letters.

It is worth knowing that most languages are not built only out of truth functional connectives. In English, it is possible to form a new sentence from any simpler sentence  $\mathcal{X}$  by saying "It is possible that  $\mathcal{X}$ ." The truth value of this new sentence does not depend directly on the truth value of  $\mathcal{X}$ . Even if  $\mathcal{X}$  is false, perhaps in some sense  $\mathcal{X}$  *could* have been true—then the new sentence would be true. Some formal languages, called *modal logics*, have an operator for possibility. In a modal logic, we could translate "It is possible that  $\mathcal{X}$ " as  $\diamond \mathcal{X}$ . However, the ability to translate sentences like these comes at a cost: The  $\diamond$  operator is not truth-functional, and so modal logics are not amenable to truth tables.

### Complete Truth Tables

In the last chapter we introduced the characteristic truth tables for the different connectives. To put them all in one place, the truth tables for the connectives of SL are repeated in Table 11.2. On the left is the truth table for negation, and on the right is the truth table for the other four connectives. Notice that the truth table for the negation is shorter than the other table. This is because there is only one metavariable here,  $\mathcal{A}$ , which can either be true or false. The other connectives involve two metavariables, which give us four possibilities of true and false. The columns to the left of the double line in these tables are called the reference columns. They just specify the truth values of the individual sentence letters. Each row of the table assigns

truth values to all the variables. Each row is thus a truth assignment—a kind of interpretation—for that sentence. Because the full table gives all the possible truth assignments for the sentence, it gives all the possible interpretations of it.

$\mathcal{A}$	$\sim \mathcal{A}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{A} \wedge \mathcal{B}$	$\mathcal{A} \vee \mathcal{B}$	$\mathcal{A} \rightarrow \mathcal{B}$	$\mathcal{A} \leftrightarrow \mathcal{B}$
T	F	T	T	T	T	T	T
F	T	T	F	F	T	F	F
		F	T	F	T	T	F
		F	F	F	F	T	T

Table 11.2: The characteristic truth tables for the connectives of SL.

The truth table of sentences that contain only one connective is given by the characteristic truth table for that connective. So the truth table for the sentence  $P \wedge Q$  looks just like the characteristic truth table for  $\wedge$ , with the sentence letters  $P$  and  $Q$  substituted in. The truth tables for more complicated sentences can simply be built up out of the truth tables for these basic sentences. Consider the sentence  $(H \wedge I) \rightarrow H$ . This sentence has two sentence letters, so we can represent all the possible truth assignments using a four line truth table. We can start by writing out all the possible combinations of true and false for  $H$  and  $I$  in the reference columns. We then copy the truth values for the sentence letters and write them underneath the letters in the sentence.

$H$	$I$	$(H \wedge I) \rightarrow H$
T	T	T
T	F	T
F	T	F
F	F	F

Now consider just one part of the sentence above, the subsentence  $H \wedge I$ . This is a conjunction  $\mathcal{A} \wedge \mathcal{B}$  with  $H$  as  $\mathcal{A}$  and with  $I$  as  $\mathcal{B}$ .  $H$  and  $I$  are both true on the first row. Since a conjunction is true when both conjuncts are true, we write a T underneath the conjunction symbol. We continue for the other three rows and get this:

		$(\mathcal{A} \wedge \mathcal{B})$				
$H$	$I$	$(H \wedge I)$	$\rightarrow$	$H$		
T	T	T	T	T		
T	F	T	F	T		
F	T	F	F	F		
F	F	F	F	F		

Next we need to fill in the final column under the conditional. The conditional is the main connective of the sentence, so the whole sentence is of the form  $\mathcal{A} \rightarrow \mathcal{B}$  with  $(H \wedge I)$  as  $\mathcal{A}$  and with  $H$  as  $\mathcal{B}$ . So to fill the final column, we just need to look at the characteristic truth table for the conditional. For the first row, the sentence  $(H \wedge I)$  is true and the sentence  $H$  is also true. The truth table for the conditional tells us this means that the whole sentence is true. Filling out the rest of the column gives us this:

		A			$\rightarrow$	B
$H$	$I$	$(H \wedge I)$			$\rightarrow$	$H$
T	T	T	T	T	T	T
T	F	T	F	F	T	T
F	T	F	F	T	T	F
F	F	F	F	F	T	F

The column of Ts underneath the conditional tells us that the sentence  $(H \wedge I) \rightarrow H$  is true regardless of the truth values of  $H$  and  $I$ . They can be true or false in any combination, and the compound sentence still comes out true. It is crucial that we have considered all of the possible combinations. If we only had a two-line truth table, we could not be sure that the sentence was not false for some other combination of truth values.

In this example, the script letters over the table have just been there to indicate how the columns get filled in. We won't need them in the final product. Also, the reference columns are redundant with the columns under the individual sentence letters, so we can eliminate those as well. Most of the time, when you see truth tables, we will just write them out this way:

$(H \wedge I)$	$\rightarrow$	$H$
T T T	T	T
T F F	T	T
F F T	T	F
F F F	T	F

The truth value of the sentence on each row is just the column underneath the *main connective* (see p. 177) of the sentence, in this case, the column underneath the conditional.

A **COMPLETE TRUTH TABLE** is a table that gives all the possible interpretations for a sentence or set of sentences in SL. It has a row for each possible assignment of T and F to all of the sentence letters. The size of the complete truth table depends on the number of different sentence letters in the table. A sentence that contains only one sentence letter requires only two rows, as in the characteristic truth table for negation. This is true even if the same letter is repeated many times, as in this sentence:

$$[(C \leftrightarrow C) \rightarrow C] \wedge \sim(C \rightarrow C).$$



The complete truth table requires only two lines because there are only two possibilities:  $C$  can be true, or it can be false. A single sentence letter can never be marked both T and F on the same row. The truth table for this sentence looks like this:

$[(C \leftrightarrow C) \rightarrow C]$	$\wedge$	$\sim$	$(C \rightarrow C)$
T	T	F	T
F	F	F	F

Looking at the column underneath the main connective, we see that the sentence is false on both rows of the table; i.e., it is false regardless of whether  $C$  is true or false.

A sentence that contains two sentence letters requires four lines for a complete truth table, as we saw above in the table for  $(H \wedge I) \rightarrow I$ .

A sentence that contains three sentence letters requires eight lines, as in this example. Here the reference columns are included so you can see how to arrange the truth values for the individual sentence letters so that all the possibilities are covered.

$M$	$N$	$P$	$M \wedge (N \vee P)$
T	T	T	T
T	T	F	T
T	F	T	T
T	F	F	F
F	T	T	T
F	T	F	F
F	F	T	T
F	F	F	F

From this table, we know that the sentence  $M \wedge (N \vee P)$  might be true or false, depending on the truth values of  $M$ ,  $N$ , and  $P$ .

A complete truth table for a sentence that contains four different sentence letters requires 16 lines. For five letters, 32 lines are required. For six letters, 64 lines, and so on. To be perfectly general: If a complete truth table has  $n$  different sentence letters, then it must have  $2^n$  rows.

By convention, the reference columns are filled in with the right most row alternating Ts and Fs. The next column over alternates sets of two Ts and two Fs. For the third column from the right, you have sets of four Ts and four Fs. This continues until you reach the leftmost column, which will always have the top half all Ts and the bottom half all Fs. This convention is completely arbitrary. There are other ways to be sure that all the possible combinations are covered, but everything is easier if we all stick to the same pattern.

### Key Terms

- |                                 |  |
|---------------------------------|--|
| 1. Complete truth table         | 7. Semantically consistent in SL           |
| 2. Interpretation               | 8. Semantically contingent in SL           |
| 3. Logical constant             | 9. Semantically logically equivalent in SL |
| 4. Nonlogical symbol            | 10. Semantically valid in SL               |
| 5. Semantic contradiction in SL | 11. Truth assignment                       |
| 6. Semantic tautology in SL     | 12. Truth-functional connective            |

## **Part V**

# **Inductive Logic and Scientific Reasoning**



## *What are Induction and Scientific Reasoning?*

### *Introduction*

Earlier in our discussion of logic we made the distinction between deductive and inductive forms of inference. Recall that while deduction and deductive inferences are inferences that attempt to guarantee the truth of the conclusion on the basis of the premises, induction and inductive inferences attempt something less ambitious. In short, induction is any form of inference in which the premises are not intended to logically entail or guarantee their conclusion, but merely to provide some support to it. In this part of the book we'll be examining inductive logic in more detail, with special attention paid to a rigorous form of inductive reasoning as applied in the sciences.

Before we begin, you may be asking yourself, "Why should we care about inductive reasoning at all?" After all, deductive inferences *guarantee* their conclusions and that's the best kind of evidential support you could hope for! Shouldn't we just spend all of our time focusing on making deductive inferences?

This is a reasonable complaint, and one worth taking seriously. While valid deduction is the absolute best form of evidential support a conclusion can receive, we should care about induction as well, for at least the following reasons.

First, many of the inferences people make every day are inductive, not deductive. Consider just the inferences you make when you go to the refrigerator to make a sandwich. Suppose you reason as follows:

(A1) Yesterday I bought sandwich ingredients and put them in my refrigerator.

(A2)  $\therefore$  So, today there are sandwich ingredients in my refrigerator.

Of course, the conclusion isn't entailed by that premise. Perhaps your roommate ate all of the cold cuts, cheese, and vegetables that you bought at the store yesterday in a late night binge. If so, then the conclusion obviously would not follow from the premise. Okay, but we can deal with

this by adding another premise, such as “No one has touched my sandwich ingredients since I put them in the refrigerator.” Still, the conclusion does not follow *deductively* from these premises since it could be that while no one has touched them the power went out in your building and all of the food in your refrigerator has gone bad!

We could go on and on like this, but the point is that most of the inferences we make are not in the form of a deductively valid argument. We often reason on limited evidence, or via less than valid forms of inference. So it is worthwhile to understand how inductive inference functions.

A second reason to care about inductive inference is that deductive inferences are fundamentally *limited*. By this I mean to emphasize the logician and philosopher C. S. Peirce’s term for inductive inference: **AMPLIATIVE LOGIC**. By ampliative, Peirce meant that the premises ‘amplify’ what can be deductively inferred from them. That is, the premises go beyond their deductive entailments.

Ampliative inference is just another word for inductive inference, but the point is the same. Inductive inferences tell us something about the world that is not already contained in the premises. For example, consider this classic deductive argument from categorical logic:

- (B1) All dogs like peanut butter.
- (B2) Rose is a dog.
- (B3)  $\therefore$  So, Rose likes peanut butter.

Now consider that if you *accept* the premises, then there’s an important sense in which the conclusion doesn’t tell you anything you didn’t already know. You haven’t learned anything about the world from this argument (assuming you already knew the premises). However, contrast that argument with the following inductive ‘version’:

- (C1) Most dogs like peanut butter.
- (C2) Rose is a dog.
- (C3)  $\therefore$  So, Rose likes peanut butter.

Obviously, the conclusion isn’t entailed by these premises (the argument isn’t valid). However, if you come to know the conclusion on the basis of these premise you *do* learn something new; namely, you learn that Rose is one of the majority of dogs who like peanut butter.

A final consideration for why induction is important is simply that most scientific inferences are made as inductive, rather than deductive, inferences. The logic of science and statistics is fundamentally inductive, not deductive. To gain a better understanding of science it is extremely important to understand how inductive inference functions.

Before moving on to the next chapters, we’ll spend the remainder of this chapter discussing some basic concepts in the history and philosophy of science in order to give some context for later discussions.

## Early Scientific Method

The history of philosophy is the history of science. Aristotle wrote lengthy treatises on the nature of mechanics and dynamics, on biology, geology, and optics. For most of history there was no sharp distinction between the scientist (a term invented by the philosopher and scientist William Whewell in 1833) and the philosopher.

One of the things that characterizes the early development of the scientific method is a commitment to **EMPIRICISM**. Empiricism is the view that the only, or perhaps the primary, source of knowledge is sensory evidence or the experience of the senses. This view is contrasted with **RATIONALISM**, which says that the source of knowledge is the internal mechanisms of reason.

Both of these views have had their philosophical defenders and both have a lot of evidence that we can acquire knowledge in the way recommended by the view. Consider, for example, proofs in geometry. The truths of mathematics in general, and geometry in particular, seem to be true no matter what. That is, these truths don't depend on making certain kinds of observations or collecting any evidence. Thus, the rationalists say, truths of this kind are derivable from pure reason. Notice that even though you and I may have made observations of a certain kind in our geometry class in order to come to know these truths, this doesn't undermine the rationalist position that such observations are not *required* in order to know geometric proofs. Strict empiricists, on the other hand, maintain that mathematical *concepts* require some kind of observation and that a being of pure reason and no senses (if such a thing is possible) would not be able to create these concepts from nothing.

Which of these views do you find more compelling? Think about where knowledge might come from and what kinds of powers are necessary in order to know things about the world. We'll focus on empiricism for now, but think about the truths of logic and their source. Are these truths the result of sensory observations or reason?

Early empiricists noticed very quickly that there are better and worse ways to observe the world around us. As a result they attempted to develop *methods* that would codify and regulate the best ways to make empirical observations and make inferences on the basis of those observations. This is an ongoing process.

There is no one "scientific method" that describes the best way to observe, predict, and model the world. Philosophers of science have been engaged in the centuries-long development of various scientific methods to help us make better inferences.

Starting with Sir Francis Bacon's *Novum Organum* and Rene Descartes' *Discourse on the Method* and ' *Principles of Philosophy* in the 17th century, philosophers of science have generally agreed that a central feature of

science is the development of a **HYPOTHESIS**. A hypothesis is simply a proposition that may or may not be true and that is subjected to some form of confirmation or disconfirmation.

Bacon's method was to begin with the simplest and most obvious observations and then use these observations to confirm or disconfirm hypotheses at greater levels of generality. Descartes, on the other hand, began by doubting the truth of everything that his senses presented to him and trying to identify hypotheses that could not possibly be doubted (such as the famous *cogito ergo sum*, or "I think, therefore I am."). Descartes then used this foundation of indubitable truths to derive more commonplace beliefs.

### *The Hypothetico-Deductive Model*

The **HYPOTHETICO-DEDUCTIVE MODEL** of scientific inference is a proposed logical form of inference that explains how scientific hypotheses can be confirmed or disconfirmed.

The hypothetico-deductive model is one of the more basic methods common to all scientific disciplines, whether it is economics, physics, or biochemistry. Its application can be divided into four stages:

1. Identify the hypothesis to be tested.
2. Generate predictions from the hypothesis.
3. Perform experiments to check whether predictions are correct.
4. If the predictions are correct, then the hypothesis is confirmed. Otherwise, the hypothesis is disconfirmed.

Suppose your portable music player fails to switch on. You might then consider the hypothesis that perhaps the batteries are dead. So you decide to test whether this is true. Given this hypothesis you predict that the music player should work properly if you replace the batteries with new ones. So you proceed to replace the batteries, which is the "experiment" for testing the prediction. If the player works again, then your hypothesis is confirmed, and so you throw away the old batteries. If the player still does not work, then the prediction is false, and the hypothesis is disconfirmed. So you might reject your original hypothesis and come up with an alternative one to test, e.g. the batteries are ok but your music player is broken.

The example above helps us illustrate a few points about science and the hypothetico-deductive method.

1. A scientific hypothesis must be testable.
2. Confirmation is not truth.
3. Disconfirmation need not be falsity.



Let's consider these points one by one.

First, notice that for us to apply the hypothetico-deductive method we must be able to test our hypothesis. The hypothetico-deductive method tells us that our hypothesis must be capable of being tested, but the method doesn't tell us (a) how to test our hypothesis or (b) how to distinguish between a testable or untestable hypothesis.

If a hypothesis cannot be tested, we cannot find evidence to show that it is probable or not. In that case it cannot be part of scientific knowledge. Consider the hypothesis that there are ghosts that have no causal efficacy, that we cannot neither see nor interact with, and which can never be detected directly or indirectly. This hypothesis is defined in such a way to exclude the possibility of being tested. It might still be true and there might be such ghosts, but we would never be in a position to know and so this cannot be a scientific hypothesis.

Second, notice that confirming the predictions of a hypothesis increases the probability that a hypothesis is correct. But in itself this does not prove conclusively that the hypothesis is correct.

To see why this is the case, we might represent our reasoning as follows:

(D1) If H then P.

(D2) P.

(D3)  $\therefore$  Therefore H.

Here H is our hypothesis "the batteries are dead", and P is the prediction "the player will function when the batteries are replaced." This pattern of reasoning is of course not valid, since there might be reasons other than H that also bring about the truth of P. For example, it might be that the original batteries are actually fine, but they were not inserted properly. Replacing the batteries would then restore the loose connection. So the fact that the prediction is true does not prove that the hypothesis is true. We need to consider alternative hypotheses and see which is more likely to be true and which provides the best explanation of the prediction. (Or we can also do more testing!)

Finally, consider that disconfirmation is sometimes not enough to demonstrate the falsity of the hypothesis. Very often a hypothesis generates a prediction only when given additional assumptions (auxiliary hypotheses). In such cases, when a prediction fails the hypothesis might still be correct.

Looking back at our example again, when we predict that the player will work again when the batteries are replaced, we are assuming that there is nothing wrong with the player. But it might turn out that this assumption is wrong. In such situations the falsity of the prediction does not logically entail the falsity of the hypothesis. We might depict the situation by this argument : ( H=The batteries are dead, A=The player is not broken.)

(E1) If both H and A, then P.

(E2) It is not the case that P.

(E3)  $\therefore$  Therefore, it is not the case that H.

This argument is of course not valid. When P is false, what follows is not that H is false, only that the conjunction of H and A is false. So there are three possibilities : (a) H is false but A is true, (b) H is true but A is false, or (c) both H and A are false. So we should argue instead :

(F1) If both H and A, then P.

(F2) It is not the case that P.

(F3)  $\therefore$  Therefore, it is not the case that both H and A are true.

Returning to our earlier example, if the player still does not work when the batteries are replaced, this does not prove conclusively that the original batteries are not dead. This tells us that when we apply the hypothetico-deductive method, we need to examine the additional assumptions that are invoked when deriving the predictions. If we are confident that the assumptions are correct, then the falsity of the prediction would be a good reason to reject the hypothesis. On the other hand, if the hypothesis we are testing has been extremely successful, then we need to be extremely cautious before we reject a hypothesis on the basis of a single false prediction. These additional assumptions used in testing a hypothesis are known as “auxiliary hypotheses.”

### *The Demarcation Criterion*

How do we distinguish between a scientific and non-scientific hypothesis? Since the philosopher of science Karl Popper, philosophers have referred to the quality or property that distinguishes between science and non-science as the **DEMARCATIION CRITERION**. According to Popper, the demarcation criterion was that a scientific hypothesis must be **FALSIFIABLE**, whereas a pseudoscientific or non-scientific hypothesis was not falsifiable.

An example of a falsifiable hypothesis is hypothesis that the Earth is flat. This hypothesis is falsifiable because it can be subjected to several tests such as measuring the angle of approaching ships, shadows, the position of the moon and its path in the sky in different parts of the Earth, and by launching a rocket and making a direct observation that the Earth is not flat. That the Earth is round is likewise falsifiable because it can be falsified by subjecting it to similar tests. The difference between these hypotheses isn't that one is “more” scientific than the other; they are equally scientific according to Popper's criterion.

However, there is one important difference between them: one is false and the other is true! The Earth is round and not flat, and we've subjected these hypotheses to many, many tests that have successfully falsified the flat Earth hypothesis and failed to falsify the round Earth hypothesis.<sup>1</sup>

<sup>1</sup> The epistemic and logical features are an interesting and growing area of philosophical and psychological research.

An example of an unfalsifiable hypothesis would be the hypothesis that reading Tarot cards will give you insight into your true purpose in life. You can't know that the hypothesis is *false*, since it is vague and relies on a specific interpretation and value system. Another non-falsifiable hypothesis is that God exists. This hypothesis is consistent with any observation, since if the hypothesis is true then it can be used to explain any outcome.

That these hypotheses are non-scientific doesn't mean they aren't meaningful, interesting, or worthwhile. They may also be true or false. What the demarcation criterion tells us is that these hypotheses aren't scientific or amenable to scientific testing. Many ethical, aesthetic, political, theological, philosophical and moral questions are non-scientific.

An important distinction, and the one that Popper initially set out to establish, is between scientific hypotheses and **PSEUDOSCIENCE**. Pseudoscientific hypotheses are not falsifiable—so, according to Popper, they are non-scientific—but they *claim to be* scientific by adopting certain practices or rhetorical moves that scientists adopt. Typically, a perfectly fine scientific hypothesis—such as the flat Earth hypothesis—will become pseudoscientific when it is not rejected appropriately. By adopting a scientific stance, its defenders can claim a kind of epistemic or logical authority that they do not have.

For example, it is pseudoscientific to defend the flat Earth hypothesis for seemingly scientific reasons while refusing to acknowledge the enormous body of evidence that the hypothesis is false. Alternatively pseudoscientific hypotheses are defended by conducting flawed experiments that do not establish what they purport to establish. Another example of a pseudoscientific hypothesis is young Earth creationism, which explains the existence of fossils and geological evidence for the age of the Earth by appeal to God or another non-falsifiable hypothesis.

One of the central philosophical challenges is that it is often unclear from the evidence available whether we should consider an ordinary scientific experiment to have falsified a hypothesis outright, or whether we should reject a different, auxiliary hypothesis. This is something that defenders of pseudoscience will often exploit. For example, antivaccination defenders will explain away the numerous falsifications of their hypothesis that vaccinations cause autism or are laced with deadly toxins by claiming that these studies do not replicate. In fact, replication of experimental results is a serious concern in scientific research and so their concern *seems* scientific. However, the studies that reject antivaccination have been extremely well established and replicated<sup>2</sup> so this concern is unwarranted. But knowing when we should reject a hypothesis is a topic of ongoing debate in the philosophy of science.

<sup>2</sup> See e.g. Taylor et. al 2014 ([link](#)) for a recent meta-analysis of this research.

### *When should we reject a hypothesis?*

When a hypothesis makes a false prediction, sometimes it can be difficult to know whether we should reject the hypothesis or whether there is something wrong with the auxiliary hypotheses. For example, astronomers in the 19th century found that Newtonian physics could not fully explain planet Mercury's orbit. It turns out that this is because Newtonian physics makes some incorrect assumptions about the structure of space and time, and you need relativity to give a more accurate prediction of the orbit. However, when astronomers discovered Uranus in 1781, they also found out that its orbit was different from the predictions of Newtonian physics. But then scientists realized that it could be explained if there was an additional planet which affected Uranus, and Neptune was subsequently discovered as a result.

Thus, in the case of Mercury's orbit the theory of Newtonian mechanics was falsified and the behavior of the orbit was only explained once we developed relativity theory. However, in the case of Uranus's orbit an auxiliary hypothesis (that another planet existed in the vicinity of Uranus) explained the deviation from the prediction of Newtonian mechanics.

In 2011, scientists in Italy reported that their experiment seemed to have shown that some subatomic particles could travel faster than the speed of light, which would seem to show that relativity is wrong. But on closer inspection, it was discovered that there was a problem with the experimental setup. So if a hypothesis has been very successful, even when a result seems to show that hypothesis is wrong, we need to make sure that the evidence is strong and reliable, and try to replicate the result and eliminate alternative explanations.

### *Causation*

Causal relationships are an extremely important part of what we know. It matters to us that our actions are efficacious. If you take medicine for an illness, it matters whether the medicine *cured* your illness or that you just happened to get better.

Philosophers of science have been interested in causation since Aristotle. Aristotle defined four kinds of causation: effective causation, material causation, formal causation, and final causation. The kind of causation that most people have in mind nowadays is what Aristotle called effective causation. That is, the effective cause of an object like a chair is the series of actions taken by a machine or carpenter to construct the chair.

By material cause, Aristotle meant the substance out of which the object is made. So, in our chair example the material cause would be the wood, nails, fabric, and so on. This is contrasted for Aristotle with the formal cause, which is the abstract form of the object or its design. Many differ-

ent chairs may share a formal cause while differing with respect to their material causes.

Finally, there is the final cause of an object. The final cause of something is the purpose or function that it was built to perform. For our chair, presumably, the final cause would be for someone to sit on.

Our philosophical and scientific understanding of causation has come a long way since Aristotle. Now we recognize that there are many conditions that need to be met in order to establish a causal hypothesis. One advancement over Aristotle's discussion of causation is the distinction between necessary and sufficient conditions.

A **NECESSARY CONDITION** for something is a state of affairs that *must* obtain in order for the effect to occur. For example, oxygen is a necessary condition for fire. The basic idea is that a necessary condition is something that is needed in order for the cause to have its expected effect. Necessary conditions are often treated as 'background conditions' because they aren't enough on their own to bring about an effect but they are, well, necessary.

A **SUFFICIENT CONDITION** is a state of affairs that is enough on its own for an effect to come about. For example, caffeine withdrawal is a sufficient condition for a headache. A sufficient condition is often treated like the 'active' cause of some effect because, given appropriate background conditions, a sufficient condition obtaining will bring about an effect.

Generally speaking, we are interested in both necessary and sufficient conditions for some effect to understand how the effect is brought about. However, we often struggle to identify the precise necessary and sufficient conditions to bring about all the effects we care about understanding. For this reason we will try to identify when two properties or states of affairs are associated.

Consider whether two properties are associated in a system that we are interested in. For example, suppose as economists we are interested in the relationship between the top marginal tax rate and the unemployment rate. Pundits and politicians often claim that wealthy people are so-called "job creators" and that raising the top marginal tax rate will cause an increase in unemployment. Suppose that their evidence is that when the top marginal tax rate in the United States was higher, unemployment was higher as well. In the United States in 1965, the top federal income tax rate was for every dollar earned over \$1.6 million was 70%.<sup>3</sup> In 2015, the top federal income tax rate for every dollar earned over approximately \$400,000 was 39.6%. Meanwhile, according to the U.S. Bureau of Labor Statistics the unemployment rate in 1965 was 4% while in 2015 it was 5%. Clearly, there is no association between the top marginal tax rate and the unemployment rate. This means that these two rates are **INDEPENDENT**. When two properties or variables are independent, knowing the value of one of them is not informative about the value of the other.

Suppose, however, that there are other reasons to think the two fea-

<sup>3</sup> This amount has been adjusted for inflation; the unadjusted amount was \$200,000.

tures are associated. An association between two features X and Y can be explained in one of four possible ways.

1. X causes Y.
2. Y causes X.
3. X and Y are both caused by a common cause, Z.
4. X and Y are coincidentally associated.

There are many associations that are merely coincidentally associated. You can see some of these associations here: [link](#).

In ?? we'll discuss causation and causal modeling in more detail. However, to understand the formalization of causation that has been happening in philosophy and machine learning in the last few decades we'll first need the underlying probability theory. That means we'll need to formalize our inductive logic.

### *Key Terms*

1. **Causation**
2. **Hypothesis**
3. **Hypothetico-deductive method**
4. **Independence**
5. **Necessary condition**
6. **Process of confirmation**
7. **Process of disconfirmation**
8. **Pseudoscience**
9. **Sufficient condition**

## Inductive Logic

Recall the how we evaluate inductive arguments from chapter 2. There we said that inductive arguments have similar evaluative standards to deductive arguments. Both kinds of argument can be evaluated for its form and for its content. We can evaluate the *form* of an inductive argument by asking whether it is **strong** or **weak**. Once we establish that an inductive argument is strong, we can ask whether its premises are true to determine if the argument is **cogent**.

You may have noticed that the definition of a strong inductive argument is somewhat vague. An inductive argument is strong just in case the premises would make the conclusion more likely, if they were true. But how *much* more likely? After all, consider the following inductive argument:

- (A1) The first number of the winning lottery ticket is 7.
- (A2) The first number of my lottery ticket is 7.
- (A3)  $\therefore$  Therefore, I have the winning lottery ticket.

Of course, these premises *do* support the conclusion somewhat. They make the conclusion more likely to be true. But is this really a strong argument? Should I *believe* or *endorse* the conclusion on the basis of these premises? Probably not.

Unfortunately there is no universally agreed-upon standard for how much evidential support some premises must provide a conclusion in order for the argument to be inductively strong. However, there is a wealth of formal inductive methods that have tried to make these evidential standards more rigorous and clear. These methods rely on the mathematical field of probability theory to provide a basis for probabilistic and inductive logics. In part ?? we'll discuss the theory of probability in much more detail, but for now let's examine the basics.

### *What is probability?*

Probability is a measure of something. Philosophers and logicians disagree about exactly *what* probability is a measure *of*, and these disagreements fall

roughly into two camps.

One group of philosophers asserts that probability is a measure of the long-term *frequency* of events. That is, if we measure some event happening or not happening over a period of time, the probability of that event is simply the ratio of the number of times it did happen divided by the number of times it *could have* happened.

Another group of philosophers asserts that probability isn't a measure of frequencies, but a measure of our *confidence* that an event will occur. That is, probability measures the degree of belief that we should adopt with respect to some proposition.

These aren't the only views that philosophers have defended, and you can probably think of other things that probability could measure. For now, we'll avoid these tricky philosophical problems and simply discuss some of the formal machinery of probability theory. Everything we say here will obtain whether you endorse one or another of these views.

In earlier chapters, we assigned a **truth value** to propositions. We said that propositions can be either true or false. In this chapter, we are going to assign probability values to propositions, but with one caveat. Each atomic proposition will describe an **ATOMIC EVENT**, by which we mean events that are mutually exclusive and exhaustive.

Consider rolling a six-sided die. Each event is one of the possible faces the die could land on. Because the die can't land on more than one side in a single roll and can't fail to land on one of the sides, these are our atomic events. We will assign a probability to each of these events using the notation " $pr(E)$ " where ' $E$ ' is shorthand for the event.

Now let's consider some of the rules that our probability values must follow.

### *The Axioms of Probability*

The axioms of probability were first expressed in this form by Soviet mathematician Andrey Kolmogorov in his book *Foundations of the Theory of Probability* in 1933. These axioms are as follows:

Non-negativity For any event  $E$ ,  $pr(E) \geq 0$

Normality  $pr(\Omega) = 1$

Finite additivity  $pr(A \vee B) = pr(A) + pr(B)$  for every  $A$  and  $B$  such that  $A \wedge B$  is false.

Let's unpack these axioms one at a time.

The first, non-negativity, says that the probability of any event must be greater than or equal to zero. In other words, all probability measures are positive numbers.



The second, normality, says that the probability of the entire event space (every possible event) is equal to one. Logicians typically use the omega symbol ( $\Omega$ ) to denote the entire event space. This has a few implications. One is that we know that *some* event in the events under consideration must occur. That is, among all of the events that we are considering, one of them is the event that will *actually* occur. Another implication of this axiom is that a probability of 1 is the maximum value that a probability measure can be. Between non-negativity and normality we know that all probability measure values are between 0 and 1, inclusive.

In the six-sided die example, the consequence of this axiom is that the probability of rolling *some* number 1-6 is equal to 1.

Finally, finite additivity is an axiom of probability which says that the probability that two mutually exclusive events,  $A$  and  $B$ , occur is simply the sum of the probabilities of each of the events. Consider our six-sided die example again. The probability of rolling a 2 or a 4 is simply the probability of rolling each of them summed together:  $\text{pr}(\text{roll two or roll four}) = \text{pr}(\text{roll two}) + \text{pr}(\text{roll four})$ .



## Mill's Methods

John Stuart Mill (1806–1873) was a British philosopher who, among other things, developed the ethical theory of Utilitarianism and made contributions to the philosophy of science. In what have come to be known as “Mill’s methods” he developed a systematized set of methods for identifying when something is a cause.

These methods are discussed in detail in Mill’s book, *A System of Logic*, which is available on the Internet Archive.<sup>1</sup>

<sup>1</sup> [Mill \(1843\) A System of Logic](#)

### *Direct method of agreement*

If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon.

– John Stuart Mill, *A System of Logic*, Vol. 1. p. 454.

For a property to be a necessary condition it must always be present if the effect is present. Since this is so, then we are interested in looking at cases where the effect is present and taking note of which properties, among those considered to be ‘possible necessary conditions’ are present and which are absent. Obviously, any properties which are absent when the effect is present cannot be necessary conditions for the effect. Consider a case from epidemiology. Imagine you are a medieval plague doctor trying to identify the cause of plague in several nearby towns. Table 14.1 lists several towns and some of their features.

The idea behind the method of agreement is that we can identify the cause of something by identifying the other property that all appearances of the cause share in common. In table 14.1, we ought to conclude that the cause of the plague is being near a port city, since that is the one thing all towns with the plague have in common.

Town Name	Near swamp?	Town well?	Port city?	Cremate dead?	Plague?
Aberdyfi	no	yes	yes	yes	yes
Berkton	yes	yes	yes	no	yes
Caelfall	no	no	yes	no	yes
Domburton	yes	no	yes	yes	yes
Eelry	yes	no	yes	no	yes
Fournemouth	yes	yes	yes	no	yes

Table 14.1: The Method of Agreement

### *Method of Difference*

If an instance in which the phenomenon under investigation occurs, and an instance in which it does not occur, have every circumstance save one in common, that one occurring only in the former; the circumstance in which alone the two instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon.

– John Stuart Mill, *A System of Logic*, Vol. 1. p. 455.

The method of difference, by contrast to the method of agreement, identifies the only difference between the situations in which the cause occurs and situations in which it doesn't occur.

Town Name	Near swamp?	Town well?	Port city?	Cremate dead?	Plague?
Quathwaite	yes	yes	yes	yes	yes
Ruthorham	no	yes	yes	yes	no
Saxondale	yes	yes	yes	yes	yes
Tarnstead	yes	yes	yes	yes	yes

Table 14.2: The Method of Difference

In table 14.2, the town of Ruthorham is the only one which isn't built near a swamp. Thus, according to the method of difference we ought to conclude that being built near a swamp is a cause of plague outbreak.

### *Joint method of agreement and difference*

If two or more instances in which the phenomenon occurs have only one circumstance in common, while two or more instances in which it does not occur have nothing in common save the absence of that circumstance; the circumstance in which alone the two sets of instances differ, is the effect, or cause, or a necessary part of the cause, of the phenomenon.

– John Stuart Mill, *A System of Logic*, Vol. 1. p. 463.

Also called simply the “joint method,” this principle simply represents the application of the methods of agreement and difference.

Town Name	Near swamp?	Town well?	Port city?	Cremate dead?	Plague?
Glenarm	yes	yes	yes	no	yes
Harmstead	yes	no	yes	yes	no
Ilfracombe	no	yes	yes	yes	yes
Jaren's Well	no	yes	yes	no	yes
Kilerth	yes	yes	yes	yes	yes

Table 14.3: The Joint Method

In table 14.3 we should conclude that the presence of a town well causes the plague.

### *Method of Residue*

Subduct from any phenomenon such part as is known by previous inductions to be the effect of certain antecedents, and the residue of the phenomenon is the effect of the remaining antecedents.

– John Stuart Mill, *A System of Logic*, Vol. 1. p. 465.

If a range of factors are believed to cause a range of phenomena, and we have matched all the factors, except one, with all the phenomena, except one, then the remaining phenomenon can be attributed to the remaining factor.

Town Name	Near swamp?	Town well?	Port city?	Swamp fever?	Giardia?	Plague?
Urmkirkey	yes	no	yes	yes	no	yes
Violl's Gar- den	yes	yes	no	yes	yes	no
Warrington	no	yes	yes	no	yes	no
Xynnar	no	yes	no	yes	yes	no
Yellowseed	no	no	yes	no	no	yes

Table 14.4: The Method of Residue

In table 14.4, suppose that we know that swamp fever is caused by being near a swamp and giardia is caused by a contaminated town well. The method of residue suggests that the only potential cause left—in this case being a port city—is the cause of plague.

### *Method of Concomitant Variation*

Whatever phenomenon varies in any manner whenever another phenomenon varies in some particular manner, is either a cause or an effect of that phenomenon, or is connected with it through some fact of causation.

– John Stuart Mill, *A System of Logic*, Vol. 1. p. 470.

If across a range of circumstances leading to a phenomenon, some property of the phenomenon varies in tandem with some factor existing in the circumstances, then the phenomenon can be associated with that factor. For instance, suppose that various samples of water, each containing both salt and lead, were found to be toxic. If the level of toxicity varied in tandem with the level of lead, one could attribute the toxicity to the presence of lead.

The method of concomitant variation can be represented as in table 14.5, where we can see that the further from a swamp the town is, the lower the population with plague. Thus, we know that distance from a swamp varies in accordance with plague so on this method we should conclude that the swamp is a cause of plague.

Town Name	Proximity to swamp	Percent of population with plague
Laencaster	111km	2%
Mirfield	40km	6%
Newsham	15km	18%
Oldham	5km	37%
Porthcrawl	2km	78%

Table 14.5: The Method of Concomitant Variation

Unlike the preceding four methods, the method of concomitant variation doesn't involve the elimination or inclusion of any features or effects. According to this method, when one property value increases or decreases this results in a change in the value of another property.

However, we know that this method (like the other of Mill's methods) is no guarantee that we've identified the cause correctly. It's possible that our potential cause is merely *associated* with the effect, rather than causing it either directly or indirectly.

Another important feature of Mill's methods is that they assume we have already determined the list of potential causes to examine. It is certainly possible that we've left the true cause off the list entirely!

### Key Terms

1. **Joint Method**
2. **Method of Agreement**
3. **Method of Concomitant Variation**
4. **Method of Difference**
5. **Method of Residue**
6. **Mill's methods**

## **Part VI**

# **Appendices**









# A

## Glossary

*Affirmative* The quality of a statement without a “not” or a “no.” 99

*Ampliative Logic* Ampliative logic is another term for inductive logic, or a logic the inferences of which do not guarantee or logically entail the truth of the conclusion on the basis of the premises. See also [Inductive Logic](#). 190, 214

*Anchoring* Anchoring occurs when we focus on the initial or most salient information we receive, even when that information is relatively unrepresentative or unhelpful. 88

*Antecedent* The sentence to the left of a conditional.. 167

*Argument* A collection of statements, called the premises, that provides evidential support to another statement, called the conclusion. 30

*Aristotelian syllogism* An argument where each statement is in one of the moods A, E, I, or O, and which has exactly three terms, arranged so that any two pairs of statements will share one term. 111

*Artificial language* A language that was consciously developed by identifiable individuals for some purpose. 55

*Atomic Event* 200

*Atomic sentence* A sentence that does not have any sentences as proper parts. 161

*Availability heuristic* The availability heuristic occurs when we only attend to the information that is readily available to us. 88

*Biconditional* A sentential connective, written as double headed arrow,  $\leftrightarrow$ , used to represent a situation where  $A$  implies  $B$  and  $B$  implies  $A$ . This is also the situation where  $A$  and  $B$  are logically equivalent. This is often expressed by the English phrase “if and only if.” 169

*Bivalent* A property of logical systems which is present when the system only has two truth values, generally “true” and “false.” 58

*Canonical form* A method for representing arguments where each premise is numbered and written on a separate line. The premises are followed by a horizontal bar and then the conclusion. Statements in the argument may be paraphrased for brevity and indicator words are removed. 31

*Categorical syllogism* An argument with two premises composed of categorical statements. 111

*Cogent* A property of arguments that holds when the argument is strong and the premises are true. 45, 199

*Cognitive bias* a habit of reasoning that can become dysfunctional in certain circumstances. Often these biases are not a matter of explicit belief. See also *fallacy* 41

*Complete truth table* A table that gives all the possible interpretations for a sentence or set of sentences in SL. 184

*Conclusion* The part of an argument that is being provided evidence. 24, 30

*Conclusion Indicator* A word or phrase such as “therefore” used to indicate that what follows is the conclusion of an argument. 31

*Conditional* The symbol  $\rightarrow$ , used to represent words and phrases that function like the English phrase “if . . . then.” 167 A kind of validity

*Conditional validity* *Confirmation bias* The tendency to discount or ignore evidence and arguments that contradict one’s current beliefs. 41, 86

*Conjunct* A sentence joined to another by a conjunction. 164

*Conjunction* The symbol  $\wedge$ , used to represent words and phrases that function like the English word “and.” 164 The sentence to the right of a

*Consequent* *Consistency* A property possessed by a set of sentences when they can all be true at the same time, but are not necessarily so. 61

*Content neutrality* the feature of the study of logic that makes it indifferent to the topic being argued about. If a method of argument is considered rational in one domain, it should be considered rational in any other domain, all other things being equal. 24

*Contingent statement* A statement that is neither a tautology nor a contradiction. 59

*Contradiction* A statement that must be false, as a matter of logic. 59

*Contradictories* Two statements that must have opposite truth values, so that one must be true and the other false. 60

*Copula* The form of the verb “to be” that links subject and predicate. 98

*Counterexample method* A method for determining whether an argument with ordinary English words for terms is valid. One consistently substitutes other English terms for the terms in the given argument to see whether one can find an argument with the same form that has true premises and a false conclusion. 136

*Critical term* the term that names things that must exist in order for a conditionally valid argument to be actually valid. 124

*Critical thinker* A person who has both sharpened their reasoning abilities using metareasoning and deploys those sharpened abilities in real world situations. 25

*Critical thinking* The use of metareasoning to improve our reasoning in practical situations. Sometimes the term is also used to refer to the results of this effort at self improvement, that is, reasoning in practical situations that has been sharpened by reflection and metareasoning. 25

*Deductive* A style of arguing where one attempts to use valid arguments. 46

*Demarcation Criterion* The demarcation criterion is the feature of a hypothesis that distinguishes it as scientific or non-scientific. 194

*Disjunct* A sentence joined to another by a disjunction. 166

*Disjunction* The symbol  $\vee$ , used to represent words and phrases that function like the English word “or” in its inclusive sense. 166 A property

*Distribution* ~~Empiricism~~ Empiricism is a view in epistemology according to which the only, or perhaps the primary, source of knowledge is sensory evidence or the experience of the senses. See also *Rationalism*. 191, 216

*Enthymeme* An argument where a premise or conclusion has been left unstated. 142

*Exclusive or* A kind of disjunction that excludes the possibility that both disjuncts are true. The exclusive or says “This or that, but not both.” 167

*Existential fallacy* A fallacy committed in an Aristotelian syllogism where the conclusion is particular but both premises are universal. 132

*Expository passage* A nonargumentative passage that organizes statements around a central theme or topic statement. 35

*Fallacy* A common mistake in reasoning. Fallacies are generally conceived of as mistake forms of inference and are generally explained by arguments represented in canonical form. See also *cognitive bias*. 43

*Fallacy of exclusive premises* A fallacy committed in an Aristotelian syllogism where both premises are negative. 130

*Fallacy of illicit process* A fallacy committed in an Aristotelian syllogism when a term is distributed in the conclusion but is not distributed in the corresponding premise. This fallacy is called “illicit major” or “illicit minor” depending on which term is not properly distributed. 129

*Fallacy of particular premises* A fallacy committed in an Aristotelian syllogism where both premises are particular. 133

*Fallacy of the undistributed middle* A fallacy committed in an Aristotelian syllogism where the middle term is not distributed in either premise. 128

*Falsifiable* A hypothesis is falsifiable if there is some evidence or observation that would demonstrate that the hypothesis is false. 194

*Formal language* An artificial language designed to bring out the logical structure of ideas and remove all the ambiguity and vagueness that plague natural languages like English. Sometimes, formal languages are also said to be languages that can be implemented by a machine. 56

*Formal logic* A way of studying logic that achieves content neutrality by replacing parts of the arguments being studied with abstract symbols. Often this will involve the construction of full formal languages. 25

*Fundamental attribution error* The phenomenon of explaining our own behavior mostly in terms of *external* influences while explaining the behavior of *others* in terms of internal characteristics. 91

*Hypothesis* A hypothesis is a proposition that may or may not be true and can be subjected to confirmation or disconfirmation. 192

*Hypothetico-Deductive Model* 192

*Inclusive or* A kind of disjunction that allows for the possibility that both disjuncts are true. The inclusive or says “This or that, or both.” 167

*Inconsistency* A property possessed by a set of sentences when they cannot all be true at the same time, but they may all be false at the same time. 61

*Independent* Two properties or variables are independent if there is no association between them. That is,  $X$  and  $Y$  are independent if  $pr(X|Y) = pr(X)$ . 197

*Inductive* A style of arguing where one attempts to use strong arguments. 46

*Inductive Logic* Inductive logic is the study of argument forms and inferences which do not guarantee or logically entail the truth of the conclusion on the basis of the premises. See also *Ampliative Logic*. 211

*Inference* the act of coming to believe a conclusion on the basis of some set of premises. 34

*Informal logic* The study of arguments given in ordinary language. 25

*Interpretation* A correspondence between nonlogical symbols of the object language and elements of some other language or logical structure. 181

*Invalid* A property of arguments that holds when the premises do not force the truth of the conclusion. The opposite of valid. 42

*Logic* The study of the structure of good inference. 24

*Logical constant* A symbol whose meaning is fixed by a formal language. Sometimes these are just called “logical symbols.” They are contrasted with NON-LOGICAL SYMBOLS. 162, 181

*Logical equivalence* A property held by a pair of sentences that must always have the same truth value. 60

*Logically structured English* English that has been regimented into a standard form to make its logical structure clear and to remove ambiguity. A stepping stone to full-fledged formal languages. 103, 108

*Main connective* The last connective that you add when you assemble a sentence using the recursive definition. 177

*Major premise* The one premise in an Aristotelian syllogism that names the major term. 112

*Major term* The term that is used as the predicate of the conclusion of an Aristotelian syllogism. 111

*Metacognition* Thought processes that are applied to other thought processes  
See also *metareasoning*. 24

*Metalanguage* The language logicians use to talk about the object language.  
In this textbook, the metalanguage is English, supplemented by certain symbols like metavariables and technical terms like “valid.” 175

*Metareasoning* Using reasoning to study reasoning. See also *metacognition*. 24

*Metavariables* A variable in the metalanguage that can represent any sentence in the object language. 175

*Middle term* The one term in an Aristotelian syllogism that does not appear in the conclusion. 112

*Minor premise* The one premise in an Aristotelian syllogism that names the minor term. 112

*Minor term* The term that is used as the subject of the conclusion of an Aristotelian syllogism. 111

*Mood-A statement* A quantified categorical statement of the form “All *S* are *P*.” 99

*Mood-E statement* A quantified categorical statement of the form “No *S* are *P*.” 99

*Mood-I statement* A quantified categorical statement of the form “Some *S* are *P*.” 99

*Mood-O statement* A quantified categorical statement of the form “Some *S* are not *P*.” 99

*Moving the goalposts* The fallacy of moving the goalposts occurs when someone claims that some piece of evidence would be sufficient for some claim *P*, only to deny that *P* once the evidence is provided. 87

*Narrative* A nonargumentative passage that describes a sequence of events or actions. 36

*Natural language* A language that develops spontaneously and learned by infants as their first language. 55

*Necessary condition* A necessary condition is a state of affairs that must obtain in order for an effect to come about. For example, oxygen is a necessary condition for fire. 172, 197

*Negation* The symbol  $\sim$ , used to represent words and phrases that function like the English word “not”. 163 The quality of a statement containing

*Negative-negative-affirmative fallacy* A fallacy committed in an Aristotelian syllogism where the conclusion is negative but both of the premises are positive or the conclusion is affirmative but one or more of the premises is negative. 131

*Nonlogical symbol* A symbol whose meaning is not fixed by a formal language.  
162, 181

*Object language* A language that is constructed and studied by logicians. In this textbook, the object part:quantificational languages are SL and QL. language is SL. 175 Ordering effects occur when the order in which we

*Ordering effect* Particular The quantity of a statement that uses the quantifier “some.” 99

*Practical argument* An argument whose conclusion is a statement that someone should do something. 30

*Predicate class* The second class named in a quantified categorical statement.  
98

*Premise* A statement in an argument that provides evidence for the conclusion.  
24, 30

*Premise Indicator* A word or phrase such as “because” used to indicate that what follows is the premise of an argument. 31

*Pseudoscience* A theory or hypothesis is pseudoscientific when it fails to meet the demarcation criterion but adopts rhetorical or argumentative practices that appear to be scientific. 195

*Quality* The status of a categorical statement as affirmative or negative. 99

*Quantified categorical statement* A statement that makes a claim about a certain quantity of the members of a class or group. 97

*Quantifier* The part of a categorical sentence that specifies a portion of a class.  
98

*Quantity* The portion of the subject class described by a categorical statement. Generally “some” or “none.” 99

*Rationalism* Rationalism is the view in epistemology that knowledge comes from internal mechanisms of reason or rationality. See also [Empiricism](#). 191, 213

*Recursive definition* A definition that defines a term by identifying base class and rules for extending that class. Also called an “inductive definition.” 177

*Rhetoric* The study of effective persuasion. 26

*Scope* The sentences that are joined by a connective. These are the sentences the connective was applied to when the sentence was assembled using a recursive definition. 177

*Semantics* The meaning of a bit of language is its meaning, including truth and falsity. 175



*Sentence letter* A single capital letter, used in SL to represent a statement. 159

*Sentence of SL* A string of symbols in SL that can be built up using according to the recursive rules given on page 176

*Sentential connective* A logical operator in SL used to combine sentence letters into larger sentences. 162

*Sentential logic* A system of logic in which statements can be defined using a recursive definition with only sentences in the base class. 177

*Simple statement of belief* A kind of nonargumentative passage where the speaker simply asserts what they believe without giving reasons. 34

*Socratic elenchus* A style of argumentation in which two statements (usually definitions) are shown to be contradictory by means of critical questioning, so an alternative is sought. 37

*Sorites categorical argument* A categorical argument with more than two premises. 147

*Sound* A property of arguments that holds if the argument is valid and has all true premises. 43

*Standard form* An Aristotelian syllogism that has been put into logically structured English with the following criteria: (1) all of the individual statements are in standard form, (2) each instance of a term is in the same format and is used in the same sense, and (3) the major premise appears first, followed by the minor premise, and then the conclusion. 112

*Standard form for a categorical statement* A categorical statement that has been put into logically structured English, with the following elements in the following order: (1) The quantifiers “all,” “some,” or “no”; (2) the subject term; (3) the copula “are” or “are not”; and (4) the predicate term. 104

*Standard form for a sorites categorical argument* A sorites argument that has been put into logically structured English with the following criteria: (1) each statement in the argument is in standard form for a categorical statement in logically structured English, (2) each instance of a term is in the same format and is used in the same sense, (3) the major premise is first in the list of premises and the minor premise is last, and (4) the middle premises are arranged so that premises that share a term are adjacent to one another. 148

*Statement* A unit of language that can be true or false. 27

*Statement mood* The classification of a categorical statement based on its quantity and quality. 99

*Strong* A property of arguments which holds when the premises, if true, mean the conclusion must be likely to be true. 45, 199

*Subject class* The first class named in a quantified categorical statement. 98

*Sufficient condition* A sufficient condition is a state of affairs that is enough on its own for an effect to come about. For example, caffeine withdrawal is a sufficient condition for a headache. [173](#), [197](#)

*Superfluous distribution rule* A rule that says that in a conditionally valid argument, the critical term will be the one that is distributed more times in the premises than is necessary to satisfy Rules 1 and 2. [153](#)

*Survivorship bias* A bias towards favoring or focusing on data collected from sources that *survived* some filter or process, rather than being eliminated by it. [90](#)

*Syllogism mood* The classification of an Aristotelian syllogism based on the moods of statements it contains. The mood is designated simply by listing the three letters for the moods of the statements in the argument, such as AAA, EAE, AII, etc. [112](#)

*Syntax* The structure of a bit of language, considered without reference to truth, falsity, or meaning. [175](#)

*Tautology* A statement that must be true, as a matter of logic. [59](#)

*Translation key* A list that assigns English phrases or sentences to variable names. Also called a “symbolization key” or simply a “dictionary.” [102](#), [114](#), [159](#)

*Truth assignment* A function that maps the sentence letters in SL onto truth values. [182](#)

*Truth evaluable* A property of some objects (such as bits of language, maps, or diagrams) that means they can be appropriately assessed as either true or false. [27](#)

*Truth value* The status of a statement with relationship to truth. For this textbook, this means the status of a statement as true or false [58](#), [181](#), [200](#)

*Truth-functional connective* an operator that builds larger sentences out of smaller ones and fixes the truth value of the resulting sentence based only on the truth value of the component sentences. [182](#)

*Unconditional validity* A kind of validity that an Aristotelian syllogism has regardless of whether the objects named by its terms actually exist. [121](#)

*Unique readability* A property of formal languages which is present when each part:quant<sub>i</sub>logicwell formed formulastatementistheproductofauniqueprocessofrecursiveconstruction.[177](#)Thequantityof fasta

*Universal validity* An argument is valid just in case its premises entail its conclusion. In other words, it is impossible for the premises to be true and the conclusion false. [40](#)

*Venn diagram* A diagram that represents categorical statements using circles that stand for classes. [101](#)

*Weak* A property of arguments that are neither valid nor strong. In a weak argument, the premises would not even make the conclusion likely, even if they were true. [45](#), [199](#)

*Zeigarnik effect* A memory bias according to which people remember the content of uncompleted tasks better than completed tasks. [92](#)



## B

### *Other Symbolic Notation*

In the history of formal logic, different symbols have been used at different times and by different authors. Often, authors were forced to use notation that their printers could typeset.

In one sense, the symbols used for various logical constants is arbitrary. There is nothing written in heaven that says that  $\sim$  must be the symbol for truth-functional negation. We might have specified a different symbol to play that part. Once we have given definitions for well-formed formulae (wff) and for truth in our logic languages, however, using  $\sim$  is no longer arbitrary. That is the symbol for negation in this textbook, and so it is the symbol for negation when writing sentences in our languages SL or QL.

This appendix presents some common symbols, so that you can recognize them if you encounter them in an article or in another book.

#### **Summary of Symbols**

negation	$\sim, \neg$
conjunction	$\&, \wedge, \cdot$
disjunction	$\vee$
conditional	$\rightarrow, \supset$
biconditional	$\leftrightarrow, \equiv$

*Negation* Two commonly used symbols are the *hoe*, ' $\neg$ ', and the *swung dash*, ' $\sim$ '. In some more advanced formal systems it is necessary to distinguish between two kinds of negation; the distinction is sometimes represented by using both ' $\neg$ ' and ' $\sim$ '.

*Disjunction* The symbol ' $\vee$ ' is typically used to symbolize inclusive disjunction.

*Conjunction* Conjunction is often symbolized with the *ampersand*, '&'. The ampersand is actually a decorative form of the Latin word 'et' which means 'and'; it is commonly used in English writing. As a symbol in a formal system, the ampersand is not the word 'and'; its meaning is given

by the formal semantics for the language. Perhaps to avoid this confusion, some systems use a different symbol for conjunction. For example, ' $\wedge$ ' is a counterpart to the symbol used for disjunction. Sometimes a single dot, ' $\cdot$ ', is used. In some older texts, there is no symbol for conjunction at all; ' $A$  and  $B$ ' is simply written ' $AB$ '.

*Material Conditional* There are two common symbols for the material conditional: the *arrow*, ' $\rightarrow$ ', and the *hook*, ' $\supset$ '.

*Material Biconditional* The *double-headed arrow*, ' $\leftrightarrow$ ', is used in systems that use the arrow to represent the material conditional. Systems that use the hook for the conditional typically use the *triple bar*, ' $\equiv$ ', for the biconditional.

*Quantifiers* The universal quantifier is typically symbolized as an upside-down A, ' $\forall$ ', and the existential quantifier as a backwards E, ' $\exists$ '. In some texts, there is no separate symbol for the universal quantifier. Instead, the variable is just written in parentheses in front of the formula that it binds. For example, 'all  $x$  are  $P$ ' is written  $(x)Px$ .

In some systems, the quantifiers are symbolized with larger versions of the symbols used for conjunction and disjunction. Although quantified expressions cannot be translated into expressions without quantifiers, there is a conceptual connection between the universal quantifier and conjunction and between the existential quantifier and disjunction. Consider the sentence  $\exists xPx$ , for example. It means that *either* the first member of the UD is a  $P$ , *or* the second one is, *or* the third one is, .... Such a system uses the symbol ' $\vee$ ' instead of ' $\exists$ '.

### Polish notation

This section briefly discusses sentential logic in Polish notation, a system of notation introduced in the late 1920s by the Polish logician Jan Łukasiewicz.

Lower case letters are used as sentence letters. The capital letter  $N$  is used for negation.  $A$  is used for disjunction,  $K$  for conjunction,  $C$  for the conditional,  $E$  for the biconditional. (' $A$ ' is for alternation, another name for logical disjunction. 'E' is for equivalence.)

In Polish notation, a binary connective is written *before* the two sentences that it connects. For example, the sentence  $A \wedge B$  of SL would be written  $Kab$  in Polish notation.

The sentences  $\sim A \rightarrow B$  and  $\sim(A \rightarrow B)$  are very different; the main logical operator of the first is the conditional, but the main connective of the second is negation. In SL, we show this by putting parentheses around the conditional in the second sentence. In Polish notation, parentheses are

notation of SL	Polish notation
$\sim$	$N$
$\wedge$	$K$
$\vee$	$A$
$\rightarrow$	$C$
$\leftrightarrow$	$E$

never required. The left-most connective is always the main connective. The first sentence would simply be written  $CNab$  and the second  $NCab$ .

This feature of Polish notation means that it is possible to evaluate sentences simply by working through the symbols from right to left. If you were constructing a truth table for  $NKab$ , for example, you would first consider the truth-values assigned to  $b$  and  $a$ , then consider their conjunction, and then negate the result. The general rule for what to evaluate next in SL is not nearly so simple. In SL, the truth table for  $\sim(A \wedge B)$  requires looking at  $A$  and  $B$ , then looking in the middle of the sentence at the conjunction, and then at the beginning of the sentence at the negation. Because the order of operations can be specified more mechanically in Polish notation, variants of Polish notation are used as the internal structure for many computer programming languages.





# C

## Quick Reference

### Characteristic Truth Tables

$\mathcal{P}$	$\sim \mathcal{P}$	$\mathcal{P}$	$\mathcal{Q}$	$\mathcal{P} \wedge \mathcal{Q}$	$\mathcal{P} \vee \mathcal{Q}$	$\mathcal{P} \rightarrow \mathcal{Q}$	$\mathcal{P} \leftrightarrow \mathcal{Q}$
T	F	T	T	T	T	T	T
T	F	T	F	F	T	F	F
F	T	F	T	F	T	T	F
F	T	F	F	F	F	T	T

### Symbolization

#### Sentential Connectives (chapter ??)

It is not the case that	$\sim P$
$P$ .	
Either $P$ , or $Q$ .	$(P \vee Q)$
Neither $P$ , nor $Q$ .	$\sim(P \vee Q)$ or $(\sim P \wedge \sim Q)$
Both $P$ , and $Q$ .	$(P \wedge Q)$
If $P$ , then $Q$ .	$(P \rightarrow Q)$
$P$ only if $Q$ .	$(P \rightarrow Q)$
$P$ if and only if $Q$ .	$(P \leftrightarrow Q)$
Unless $P$ , $Q$ .	$(P \vee Q)$
$P$ unless $Q$ .	$(P \vee Q)$

#### Predicates (chapter ??)

All $F$ s are $G$ s.	$\forall x(Fx \rightarrow Gx)$
Some $F$ s are $G$ s.	$\exists x(Fx \wedge Gx)$
Not all $F$ s are $G$ s.	$\sim \forall x(Fx \rightarrow Gx)$ or $\exists x(Fx \wedge \sim Gx)$
No $F$ s are $G$ s.	$\forall x(Fx \rightarrow \sim Gx)$ or $\sim \exists x(Fx \wedge Gx)$

*Identity (section ??)*

- Only  $j$  is  $G$ .  $\forall x(Gx \leftrightarrow x = j)$   
 Everything besides  $j$  is  $G$ .  $\forall x(x \neq j \rightarrow Gx)$   
 $j$  is more  $R$  than anyone else.  $\forall x(x \neq j \rightarrow Rjx)$   
 The  $F$  is  $G$ .  $\exists x(Fx \wedge \forall y(Fy \rightarrow x = y) \wedge Gx)$

*'The  $F$  is not  $G$ ' can be translated two ways:*

- It is not the case that the  $F$  is  $G$ .  $\sim \exists x(Fx \wedge \forall y(Fy \rightarrow x = y) \wedge Gx)$   
 (wide)  
 $\exists x(Fx \wedge \forall y(Fy \rightarrow x = y) \wedge \sim Gx)$   
 The  $F$  is non- $G$ . (narrow)

*Using identity to symbolize quantities*

*There are at least  $n$   $F$ s.*

- one**  $\exists xFx$   
**two**  $\exists x_1 \exists x_2 (Fx_1 \wedge Fx_2 \wedge x_1 \neq x_2)$   
**three**  $\exists x_1 \exists x_2 \exists x_3 (Fx_1 \wedge Fx_2 \wedge Fx_3 \wedge x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_2 \neq x_3)$   
**four**  $\exists x_1 \exists x_2 \exists x_3 \exists x_4 (Fx_1 \wedge Fx_2 \wedge Fx_3 \wedge Fx_4 \wedge x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_1 \neq x_4 \wedge x_2 \neq x_3 \wedge x_2 \neq x_4 \wedge x_3 \neq x_4)$   
 **$n$**   $\exists x_1 \cdots \exists x_n (Fx_1 \wedge \cdots \wedge Fx_n \wedge x_1 \neq x_2 \wedge \cdots \wedge x_{n-1} \neq x_n)$

*There are at most  $n$   $F$ s.*

One way to say 'at most  $n$  things are  $F$ ' is to put a negation sign in front of one of the symbolizations above and say  $\sim$  'at least  $n + 1$  things are  $F$ .' Equivalently:

- one**  $\forall x_1 \forall x_2 [(Fx_1 \wedge Fx_2) \rightarrow x_1 = x_2]$   
**two**  $\forall x_1 \forall x_2 \forall x_3 [(Fx_1 \wedge Fx_2 \wedge Fx_3) \rightarrow (x_1 = x_2 \vee x_1 = x_3 \vee x_2 = x_3)]$   
**three**  $\forall x_1 \forall x_2 \forall x_3 \forall x_4 [(Fx_1 \wedge Fx_2 \wedge Fx_3 \wedge Fx_4) \rightarrow (x_1 = x_2 \vee x_1 = x_3 \vee x_1 = x_4 \vee x_2 = x_3 \vee x_2 = x_4 \vee x_3 = x_4)]$   
 **$n$**   $\forall x_1 \cdots \forall x_{n+1} [(Fx_1 \wedge \cdots \wedge Fx_{n+1}) \rightarrow (x_1 = x_2 \vee \cdots \vee x_n = x_{n+1})]$

*There are exactly  $n$   $F$ s.*

One way to say ‘exactly  $n$  things are  $F$ ’ is to conjoin two of the symbolizations above and say ‘at least  $n$  things are  $F$ ’  $\wedge$  ‘at most  $n$  things are  $F$ .’ The following equivalent formulae are shorter:

- zero**  $\forall x \sim Fx$
- one**  $\exists x [Fx \wedge \sim \exists y (Fy \wedge x \neq y)]$
- two**  $\exists x_1 \exists x_2 [Fx_1 \wedge Fx_2 \wedge x_1 \neq x_2 \wedge \sim \exists y (Fy \wedge y \neq x_1 \wedge y \neq x_2)]$
- three**  $\exists x_1 \exists x_2 \exists x_3 [Fx_1 \wedge Fx_2 \wedge Fx_3 \wedge x_1 \neq x_2 \wedge x_1 \neq x_3 \wedge x_2 \neq x_3 \wedge \sim \exists y (Fy \wedge y \neq x_1 \wedge y \neq x_2 \wedge y \neq x_3)]$
- $n$**   $\exists x_1 \cdots \exists x_n [Fx_1 \wedge \cdots \wedge Fx_n \wedge x_1 \neq x_2 \wedge \cdots \wedge x_{n-1} \neq x_n \wedge \sim \exists y (Fy \wedge y \neq x_1 \wedge \cdots \wedge y \neq x_n)]$

*Specifying the size of the UD*

Removing  $F$  from the symbolizations above produces sentences that talk about the size of the UD. For instance, ‘there are at least 2 things (in the UD)’ may be symbolized as  $\exists x \exists y (x \neq y)$ .

### *About the authors*

*P. D. Magnus* is an associate professor of philosophy in Albany, New York. His primary research is in the philosophy of science, concerned especially with the underdetermination of theory by data.

*J. Robert Loftis* is an associate professor of philosophy at Lorain County Community College in Elyria, Ohio. He received his Ph.D. in philosophy from Northwestern University.

*Cathal Woods* is Batten Associate Professor of Philosophy at Virginia Wesleyan University. He received his Ph.D. in philosophy from The Ohio State University.

*Adam Edwards* is a graduate student at the University of Illinois at Urbana-Champaign.

