HANS HALVORSON

HOW LOGIC WORKS

FORTHCOMING FROM PRINCETON UNIVERSITY PRESS

Contents

Preface

This book has no subject matter — or, to be more precise, it's about both everything and nothing. For the science of logic has no doctrines or creeds. There is no set of beliefs that distinguishes the logical people from the non-logical people, not the beliefs of the European Enlightenment, nor the deliverances of contemporary natural science, nor the opinions of some Princeton philosopher. Simply put, learning the science of logic can't be reduced to learning any particular facts at all; it's learning a skill, namely, the skill to discern between good and bad arguments.

There are numerous reasons why you need this skill, no matter what you end up doing with your life. First, being logical will help you reason about how to get what you want. Second, many of our society's best jobs require strong logic skills — whether it be programming computers, buying stock options, curing diseases, prosecuting criminals, discovering alternative energy sources, or interpreting the constitution. (And that's not even to speak of that all-important task of parenting and raising intellectually healthy children.) Third, regardless of your ideological bent — whether you're religious, atheist, or agnostic — you surely want to do everything you can to ensure that your view aligns with reality. For this task, logic is an invaluable tool, if only to protect you from the plethora of bad arguments you'll hear in your life.

Our goal here is no more or less than to initiate you into the most up-to-date account of what makes an argument good. We'll give you the tools, but it's up to you to decide how you're going to use them.

Note to the instructor, or the person indecisive about which logic book to use: To express a thought, or make an argument, or to formulate a theory, you have to pick a language to use. Similarly, to become more logical, you have to adopt some particular system of logic. I've made a choice for you here, and here is my rationale.

The first choice point is between "trees" and "arguments." The advantage of trees is that they are really easy, and require little mental exertion. But wait, isn't the goal to become the most excellent logical thinkers that we can be? People don't become better logical

thinkers by being taught a recipe that somebody else found, and that lets them effectively turn their minds off. But that's precisely the point of logical trees: they give students a *recipe* for evaluating simple arguments. If the goal were to manufacture an army of logical automata, then I might well use trees. But since I'm teaching human beings, I prefer to teach them that distinctively human skill of making and evaluating arguments. Accordingly, this book focuses primarily on how to make rigorous arguments, and secondarily on the (non-algorithmic) skill of detecting bad arguments.

The second choice point is whether to represent arguments using a "Fitch-style" or a "Lemmon-style" system. This is a difficult decision. Fitch-style is very intuitive and has a shallow learning curve. Unfortunately, Fitch-style argument is opaque to reflection: it's difficult for students to see why Fitch-style works, and even more difficult for them to imagine how the rules might have been different (thereby reinforcing an unfortunate perception that there is no human creativity involved in formulating the laws of logic). In contrast, Lemmon-style argument has a steeper learning curve, but it's more flexible and transparent to reflection. It's relatively easy to see why the system works, and it's easy to tweak the rules and see how things would come out different. So, the steeper learning curve of the Lemmon approach is a price I'm willing to pay to enable my reader to become a more clear and creative logical thinker.

Note to the reader: We put an asterisk to the left of exercises that might be perceived as more difficult than the preceding ones. In the case of proofs, it often occurs that a difficult result is basically a version of the law of excluded middle: $P \vee \neg P$. So, if you get stuck, you might try first to prove excluded middle, and then to use that to get the result you want.

1
Logic for Humans

You're a curious person, I suspect. You probably already flipped through the pages of this book, in which case you may have run across some unfamiliar symbols. You might have found yourself intrigued — like an archaeologist discovering ancient runes. Or you might have been put off — thinking that this book is for quantitative people.

That's what I assumed at first. I wanted to spend the days of my life thinking about the big questions of human existence — what exists, what we can know, and how we should live. Calculate the derivative of a function? Solve a differential equation? No thank you. I'll leave that to the people who want to build better bridges. I'd prefer to move on to the really meaningful and enriching topics.

But I discovered that it's a false dilemma. In fact, it's not a dilemma at all. Symbolic logic is not only for mathematics, and it's by no means a diversion from the really deep questions of human life. In fact, symbolic logic represents the best account we have of what it means to be *rational*.

Although logic is symbolic, it's not really "mathematical" in any sense that puts it in opposition to humanistic endeavors (such as literature, poetry, history, philosophy, etc.). Yes, mathematics is a human activity that displays logical thinking in a particularly clear way. But logic itself is involved in any type of human thinking that aims at finding the truth. If you've ever argued for a claim, or evaluated someone else's argument, then you were using logic — whether you realized it or not, and whether or not you did a good job of it.

The goal of this book is simple: it's to make you conscious of how you already use logic, and thereby to become even better at it. If you learn to do symbolic logic, then you will become a better thinker, and you will understand better what it means to be a good thinker.

Arguments

Many logic books begin by saying: "The subject matter of logic is . . . " I think these statements are always a bit misleading. In one sense, logic doesn't have a subject matter at all. Logic isn't *about* something, it's a way of life.

Let's begin by trying to see ourselves from the outside. Just imagine that you are an alien who has landed on earth, and you're trying to understand what human beings are doing when they say that they are thinking logically. Imagine that there are two people, say Anne and Bernt, and that Anne is trying to convince Bernt that something is true. Anne might proceed as follows:

> Of course gay marriage should be legal. Only people with some backwards religious view would believe otherwise.

Here Anne is trying to convince Bernt that gay marriage should be legal. But she doesn't try to coerce him with physical force, or even with intellectual intimidation. Instead, she offers Bernt a *reason* why he should accept her conclusion. To be more clear, the **conclusion** of Anne's argument is the statement, "Gay marriage should be legal." The reason that Anne gives for this conclusion — "Only people with some backwards religious view believe otherwise" — will be called the **premise** of the argument. Thus, the argument consists of a premise, and a conclusion that is supposed to be supported by the premise.

Thus, we have three things in play: **argument**, **conclusion**, and **premise**. The argument itself is made up of the conclusion and the premise. The conclusion and premise themselves are particular sentences. Notice, moreover, that these sentences are *assertions*, i.e. they make a statement that is either true or false. Thus, an argument is built out of assertions (or statements), some of which are premises, and one of which is the conclusion.

The key point about an argument is that it's more than just a disconnected collection of statements. Suppose that I have ten notecards, each of which has a statement on it. If I shuffle them up, and hand them to you, then I haven't given you an **argument**. For a collection of statements to be an argument, there has to be some implied sense in which some of the statements stand in a special relation to another one of the statements. In the notecard analogy, I'd have to hand you a first batch of notecards and say, "these are my premises," and then I'd have to hand you another notecard and say, "and this statement is my conclusion — which, I claim, follows logically from those premises."

What is this relation of "following logically" that I claim holds between my premises and my conclusion? We all know it when we

see it, and we have many words for it — words such as "supports" or "implies" or "entails" or "shows that" or "grounds". That is, we say things like, "The fact that there are cookie crumbs on the carpet *shows that* my son was eating in the living room."

It's this relation — whatever it is — that we really want to understand. We want to know: when does this relation hold between statements? When does one statement imply another? There is simply nothing more basic to human rationality than the notion of one statement implying another.

We will make a lot of progress in clarifying the notion of implication. But we're not going to make progress by means of a head-on assault. That is, we're not going to offer you a definition of the form:

> To say that one statement implies another means that . . .

Such a definition would be interesting, but it's not what this book is about. This book is more of a training manual for logic connoisseurs. Just as a wine connoisseur knows a good wine when he tastes one, so a logic connoisseur knows a good argument when she sees one.

Logical form

The study of logic began in ancient Greece — and possibly in other places at other times, although that history is less well known to us. It all began with a single insight, which you've probably already had yourself. This insight is that whether or not an argument is good depends only on its **form**, and not on its **content**. To explain this distinction, we need to back up for a second and explain what we mean by saying that an argument is "good." Consider the following argument:

> All whales are mammals.
> David Hasselhof is a whale.
> Therefore, David Hasselhof is a mammal.

Here there are two premises, and one conclusion. We've used the word, "therefore" to indicate what the conclusion is. But in truth, the word "therefore" isn't part of the content of the conclusion. The conclusion is just the proposition, "David Hasselhof is a mammal."

Is this a good argument? I hope that your answer is, "it depends." It certainly isn't a perfect argument, because it involves a false statement, namely that David Hasselhof is a whale. Or maybe you don't know anything about David Hasselhof? (Such deplorable lack of cultural knowledge these days!) Suppose that David Hasselhof were actually a famous whale in a book by an obscure author named Melvin Hermanville. In that case — if Hasselhof were a whale — then would it be a good argument? Yes, it would definitely be a good argument.

If you're a philosophy type, then you might still be doubtful. You might be thinking, "it all depends on what you mean by 'good'." If by "good" we mean "interesting, informative, and non-trivial," then that argument might not be very good. However, logic has no use for subjective words such as "interesting." Logic is the *science* of good arguments, and it's interested in isolating an *objective* sense of goodness in arguments.

The insight — passed on to us by the ancient Greeks — is that we can define "good argument" in an objective sense by factorizing goodness into two distinct pieces. The first of the two pieces is easy to understand, but difficult to agree upon in practice: are the premises true? The second piece is a bit more elusive, but forms the subject matter of logic as an objective science: do the premises support (or imply, or entail) the conclusion? If the premises do imply the conclusion, then we say that the argument is valid.

Definition. An argument is said to be **valid** if its premises imply its conclusion.

The notion of validity isn't concerned with whether the premises or conclusion are true or false. The question, instead, is a conditional one: *if* the premises were true, *then* would the conclusion be true?

You should be able to think of cases where you would agree that the premises support the conclusion, even though you think that the premises are false. It might help to use the phrase, "the premises *would* support the conclusion", the idea being that if they *were* true, then they would imply that the conclusion is also true.

You should also be able to think of arguments where the premises *and* conclusion are true, but the premises do *not* imply the conclusion. For example, the following is a true premise: "I love coffee." The following is also true: "I am over six feet tall." But to make an argument from my loving coffee to my above average height would be patently invalid. Logical validity is all about the connection between premises and conclusion; it's not directly concerned with the question of whether the premises or conclusion are true.

Sameness of form

How do we get our hands on this elusive notion of validity, and the related notion of implication? Let's begin by looking at obvious cases — where an argument is obviously valid, or obviously invalid. For example, the argument above was obviously valid. But the argument below is obviously invalid:

> Princeton is a town in New Jersey.
> Therefore, God doesn't exist.

Now, you might actually think that both of these statements are true. But that most certainly doesn't mean that the first statement implies the second. Some true statements just don't have anything to do with each other. And that's why this argument is invalid — because the premise doesn't give the right kind of support for the conclusion.

Consider another argument:

> All whales are predators.
> Bambi is a whale.
> Therefore, Bambi is a predator.

Is that a good argument? Before you answer, remember that validity doesn't have anything to do with whether you believe the premises or the conclusion. It's merely a matter of whether there is the right kind of connection between premise and conclusion.

Imagine for a moment that you just learned English, and that you aren't yet familiar with the word "whale", or with the name "Bambi." For all you know, "whale" might mean the same thing as "tiger". And for all you know, "Bambi" might be the name of a tiger at the Philadelphia zoo.

Here's the amazing thing: you don't have to know anything about the meaning of the words "whale", "predator" and "Bambi" to know that this argument is valid. How do you know it's valid? I'm not going to try to answer that question directly. I'm going to assume that you share my intuition that it is obviously valid. If you're still not convinced, let me put it this way:

> If all whales were predators, and if Bambi were a whale, then would it follow that Bambi is a predator?

Now it seems pretty obvious, doesn't it?

We said that the validity of that argument doesn't depend at all on what the "content words" mean. In other words, if an argument is valid, then it should remain valid no matter how we interpret the content words, or even if we replace the content words with different ones. Thus, given a valid argument (such as the one above), we should be able to create at sort of "mad lib argument" with variables that can be filled in by content words.

> All X are Y.
> m is an X.
> Therefore, m is a Y.

No matter what words you put in for X, Y, and m (provided that the result is a well formed sentence), you get a valid argument.

The thing above with the variables, it's like a blueprint for constructing arguments. Choose some content words, plug them in, and

ta da, you have a valid argument. Let's call it an **argument form**. In this case, it's a valid argument form, because no matter what words you plug in, the argument comes out as valid.

But how did we know that those arguments were valid in the first place? To be honest, it's just our intuition that tells us that these arguments are valid. Nobody found a tablet of stone on a mountain with the argument form above. Instead, that argument form was written down by a human being in an effort to capture what is common in a bunch of arguments that we feel (intuitively) to be valid.

That's how we'll proceed in the first part of this book: we will collect several basic argument forms that seem obviously valid. Then we'll learn how to string valid argument forms together to create longer valid arguments.

Glossary

antecedent The antecedent of a conditional is the sentence that occurs after "if", e.g. in $\phi \to \psi$ the antecedent is ϕ 19

atomic sentences have no internal structure, in particular, contain no logical connectives. 48

conditional A conditional sentence is one whose main connective is $\to$ 19

consequent The consequent of a conditional is the sentence that occurs after "then", e.g. in $\phi \to \psi$ the consequent is ψ 19

counterexample A counterexample to an argument form is an interpretation relative to which the premises of the argument are true, and the conclusion of the argument is false 20, 64

disjunction A disjunction is a sentence whose main connective is $\vee$ 16

main column is the column in the truth table of a sentence corresponding to the main connective of that sentence 61

model A model of a theory is an interpretation in which all the theory's sentences are true. 148

reconstrual A reconstrual is a map from atomic sentences to sentences. 48

sequent A sequent consists of n sentences, a turnstile, and another sentence. It represents the symbolic form of an argument 16

substitution instance A substitution instance of a sentence is any other sentence that results from the first by a uniform replacement of non-logical terms. i.e. it is any sentence that could result from translating that sentence to another language. 47

translation A translation is a map from sentences to sentences, recursively defined in terms of a reconstrual. 48

valid An argument is valid just in case its premises provide decisive support for its conclusion 10