Sentential Logic Primer

Richard Grandy
Rice University[1]

Daniel Osherson
Princeton University[2]

July 23, 2004

[1]rgrandy@rice.edu
[2]osherson@princeton.edu

Copyright

Preface

Students often study logic on the assumption that it provides a normative guide to reasoning in English. In particular, they are taught to associate connectives like "and" with counterparts in Sentential Logic. English conditionals go over to formulas with $\rightarrow$ as principal connective. The well-known difficulties that arise from such translation are not emphasized. The result is the conviction that ordinary reasoning is faulty when discordant with the usual representation in standard logic. Psychologists are particularly susceptible to this attitude.

The present book is an introduction to Sentential Logic that attempts to situate the formalism within the larger theory of rational inference carried out in natural language. After presentation of Sentential Logic, we consider its mapping onto English, notably, constructions involving "if ... then ...." Our goal is to deepen appreciation of the issues surrounding such constructions.

We make the book available, for free, on line (at least for now). Please be respectful of the integrity of the text. Large portions should not be incorporated into other works without permission.

Feedback will be greatly appreciated. Errors, obscurity, or other defects can be brought to our attention via rgrandy@rice.edu or osherson@princeton.edu. The provenance of revisions will be acknowledged as new versions are produced.

Richard Grandy
Daniel Osherson

Note to students

So . . . We're going to do some logic together, is that it?

OK. We're on board. We'll do our best to be clear. Please forgive us if we occasionally let you down (wandering into impenetrable prose). In that case, don't hesitate to send us (polite) email. We'll try to fix the offending passage, and send you back a fresh (electronic) copy of the book.

Now what about you? What can we expect in return? All we ask (but it's a lot) is that you be an *active learner*. Yes, we know. Years of enforced passivity in school has made education seem like the movies (or worse, television). You settle back in your chair and let the show wash over you. But that won't work this time. Logic isn't so easy. The only hope for understanding it is to read slowly and grapple with every idea. If you don't understand something, you must make an additional effort before moving on. That means re-reading the part that's troubling you, studying an example, or working one of the exercises. If you're reading this book with someone else (e.g., an instructor), you should raise difficulties with her *as they arise* rather than all-at-once at the end. Most important, when the discussion refers to a fact or definition that appears earlier in the book, *go back and look at it* to make sure that things are clear. Don't just plod on with only a vague idea about the earlier material. To facilitate this process, read the book with a note pad to hand. When you go back to earlier material, jot down your current page so you can return easily.

Now's the time to tell you (while we're still friends) that a normal person can actually come to enjoy logic. It looks like logic is about formulas in some esoteric language, but really it's about people. But you won't believe us until we've made significant progress in our journey. So let's get going, if you have the courage. Perhaps we'll meet again in Chapter 1.

Note to instructors

The present text differs from most other logic books we know in the following ways.

(a) Only the sentential calculus is treated.

(b) Sentential semantics are built around the concept of *meaning* (sets of truth-assignments).

(c) The derivation system is particularly simple in two respects. Assumptions are cancelled by filling in open circles that flag live hypotheses; also, there is only one rule that cancels assumptions.

(d) It is shown how probabilities can be attached to formulas.

(e) Sentential Logic is evaluated as a theory of "secure inference" in English.

(f) Having noted deficiencies in Logic's treatment of English conditionals, several alternatives to standard logic are explored in detail.

There are some exercises, but not enough. We will gratefully acknowledge any assistance in this matter (contact us about format).

Acknowledgements

Whatever is original in our discussion is little more than reassembly of ideas already developed by other scholars. The earlier work we've relied upon is acknowledged along the way.

The book has benefitted from perspicacious feedback from Michael McDermott, and from eagle-eye proofing by Roger Moseley (a surgeon!). These gentlepeople should not be held responsible for errors and confusions that remain.

The pictures that grace the chapters were composed by Anne Osherson.

We acknowledge support from NSF grant IIS-9978135 to Osherson.

Chapter 1

Introduction

1.1 Reasoning

Suppose you had to choose one feature of mental life that sets our species apart from all others. It should be a capacity exercised virtually every day that affects the human condition in countless ways. Although present in attenuated form in other mammals, it should reach its most perfected state in people. What feature of mental life would you choose?

It seems obvious that the only contender for such a special human capacity is *love*. What could be more remarkable about our species than the tendency of its members to form stable and deeply felt attachments to each other, transcending generations and gender, often extending to entire communities of heterogeneous individuals? Love is surely what separates us from the grim world of beasts, and renders us worthy of notice and affection. (For discussion, see [49].)

Alas, this book is about something else. It concerns *reason*, which is also pretty interesting (although not as interesting as love). The capacity for reasoning may be considered the second most distinguishing characteristic of our species. We're not bad at it (better, at any rate, than the brutes), and like love it seems necessary to keep the human species in business.

To make it clearer what our subject matter is about, let us consider an example of reasoning. Suppose that c_1 and c_2 are cannonballs dropped simultaneously from the top story of the Tower of Pisa. They have identical volumes, but c_1 weighs one pound whereas c_2 weighs two. Could c_2 hit the ground before c_1? If it does, this is almost surely due to the fact that 2-pound objects fall faster than 1-pound objects. Now, c_2 can be conceived as the attachment of two, 1-pound cannonballs, so if it fell faster than c_1 this would show that two, 1-pound objects fall faster attached than separately. This seems so unlikely that we are led to conclude that c_2 and c_1 will land at the same time. Just this line of thought went through the mind of Galileo Galilei (1564 - 1642), and he probably never got around to dropping cannonballs for verification (see Cooper [21]).

Galileo's thinking, as we have presented it, has some gaps. For one thing, there is unclarity about the shapes of the two 1-pound cannonballs that compose c_2. Still, the reasoning has evident virtue, and we are led to wonder about the biological conditions necessary for an organism to dream up such clever arguments, or even to follow them when presented by someone else.

Related questions arise when thought goes awry. In an often cited passage from a prestigious medical journal, the author infers the probability of breast cancer given a negative mammogram from nothing more than the probability of a negative mammogram given breast cancer, taking them to be equal. That no such equivalence holds in general is seen by comparing the high probability of having swum in the ocean given you live in the Bahamas with the low probability of living in the Bahamas given you've swum in the ocean (or comparing the probability of an animal speaking English given it's a mammal with the probability of it being a mammal given it speaks English, etc.). What is it about our brains that allow such errors to arise so easily?

The causes and consequences of good and bad thinking have been on the minds of reflective people for quite a while. The Old Testament devotes space to King Solomon's son and successor Rehoboam. Faced with popular unrest stemming from his father's reliance on forced labor, Rehoboam had the stunning idea that he could restore order with the declaration: "Whereas my father laid upon you a heavy yoke, I will add to your yoke. Whereas my father chastised you with whips, I shall chastise you with scorpions." (Kings I.12.11) Social disaster ensued.[1] Much Greek philosophical training focussed on distinguishing reliable forms of inference from such gems as:

> This mutt is your dog, and he is the father of those puppies. So, he is yours and a father, hence *your father* (and the puppies are your siblings).[2]

The authors of the 17th century textbook *Logic or the Art of Thinking* lament: "Everywhere we encounter nothing but faulty minds, who have practically no ability to discern the truth. . . . This is why there are no absurdities so unacceptable that they do not find approval. Anyone who sets out to trick the world

[1]Other memorable moments in the history of political miscalculation are described in Tuchman [104].

[2]See Russell [86, I.X].

is sure to find people who will be happy to be tricked, and the most ridiculous idiocies always encounter minds suited to them."[3] [8, pp. 5-6]

That reasoning sometimes goes astray is one motive for trying to reveal its psychological and neurological origins. With knowledge of mechanism, we would be in a better position to improve our thinking, train it for special purposes, and fix it when it's broken by disease or accident. The scientific study of human reasoning has in fact produced a voluminous literature. (See [9, 34, 46] for introductions.)

1.2 Orientation and focus

The goal of the present work is to present aspects of the formal theory of logic that are pertinent to the empirical study of reasoning by humans. Logic helps to define the maximum competence that an ideal reasoning agent can hope to achieve, just as the nature of electromagnetic radiation defines the maximum information that the eye can derive from light. In both cases, knowledge of such limits seems essential to understanding the strengths and weaknesses of the human system, and ultimately to uncovering the mechanisms that animate it.

The concern for optimality makes logic into a *normative* discipline inasmuch as it attempts to characterize how people *ought* to reason. In contrast, psychological theories are *descriptive* inasmuch as they attempt to characterize how people (or a given class of persons) *actually* reason. We'll see, however, that the boundary between normative and descriptive is not always easy to trace. Suppose we tell you:

(1) If you negate the negation of a claim then you are committed to the claim without negation.

We have in mind inferences from a claim like (2)a, below, to (2)b.

(2) (a) It's not true that John does not have anchovies in his ice cream.

[3] An introduction to the history of reflection on reasoning is provided in Glymour [35].

(b) John has anchovies in his ice cream.

This example makes (1) look good. But suppose your dialect includes what look like double negations, such as:

John ain't got no anchovies in his ice cream.

Then there seems to be no commitment to (2)b, and (1) appears ill-founded. The status of (1) seems therefore to depend on how negation functions in your language, which is ultimately a descriptive fact about your psychology. Normative and descriptive considerations become especially entangled when *conditional sentences* are at issue like "*If* John has anchovies in his ice cream *then* his girl friend will call it quits." (Conditionals take center stage in later chapters.)

We shall therefore present a basic formalism used in normative theories of inference, but we will also discuss issues that arise in attempting to interpret the formalism as advice about good thinking. You'll see that when ordinary reasoning conflicts with the dictates of standard logic, it's sometimes not clear which is to blame. Does the reasoning fail to meet some legitimate criterion of rationality? Or does the logic fail to be relevant to the kind of reasoning at issue? Answering this (or at least being sensitive to its nuances) is essential to understanding the character of human thought.

The authors might be wrong, of course, in suggesting that normative analysis is a precondition to descriptive insights about reasoning. It is possible that descriptive theorists would make more progress by just getting on with things, that is, by turning their backs to the philosophers' formalisms, and staying focussed on the empirical issues of actual reasoning. You'll have to decide whether you trust us enough to plod through the forbidding chapters that follow this one, and to face up to the exercises. If not, you are free to return to topics more purely psychological — for example, to Carl Jung's theory of archetypes in the collective unconscious [56]. Enjoy!

Some more remarks on our topic may be worthwhile (if you're still there). A broader view of reason would place optimal decision-making and planning within its purview, not just the optimal use of information for reaching new beliefs. Questions about decision-making often take the form: given that you believe such-and-such, and you desire so-and-so, what is it rational to *do*? We will not address this kind of question, since our narrower perspective will keep us plenty busy. We will be concerned only with the information that can be gleaned from sentences, not with the choices or actions that might be justified thereby. Indeed, we must be content in this book with presenting only a small fraction of contemporary logic in this restricted sense, limiting ourselves to the most elementary topics. Even in this circumscribed realm, various technical matters will be skirted, in favor of discussing the bearing of logical theory on human thought.[4]

It will likewise be expedient to neglect foundational issues about logic, e.g., concerning the metaphysical status of validity, probability, and truth itself. Thus, our starting point is that truth and falsity can be meaningfully attributed to certain linguistic objects, called "sentences." Most of the beguiling questions raised by such an innocent beginning will be left to other works.[5]

1.3 Reason fractionated

Contemporary logic embraces a tremendous variety of perspectives on reasoning and language. To organize the small subset of topics to be examined in what follows, we distinguish three kinds of logic. They may be labeled *deductive, inductive*, and *abductive*. These are traditional terms, each having already been employed in diverse ways. So let us indicate what we have in mind by the three rubrics.

Deductive logic is the study of *secure inference*. From the sentences

> For every bacterium there is an antibiotic that kills it.
>
> No antibiotic kills Voxis.

the further sentence

> Voxis is not a bacterium.

[4] For an introduction to decision and choice, see Resnik [83] and the more advanced Jeffrey [53].

[5] The issues are masterfully surveyed in Kirkham [60] and in Soames [92].

may be securely inferred. The security consists in the fact that the former sentences can be true only if the latter one is too; that is, it is impossible for the premises to be true yet the conclusion be false. Deductive logic attempts to give precise meaning to this idea. We will present *sentential* logic, which is designed to elucidate secure inferences that depend on some simple features of sentences.

Inductive logic concerns *degrees of belief*. In the formalism we will be studying, degrees of belief are quantified as *probabilities*, hence numbers from the interval 0 to 1. These are attached to sentences, subject to various restrictions. For example, the probabilities that belong to

> Voxis is a bacterium and it is deadly.
>
> Voxis is a bacterium and it is not deadly.

must sum to the probability that goes with: Voxis is a bacterium.

Abductive logic (as we use this term) bears on the conditions under which sentences should be *accepted*. In contrast to degrees of belief as represented by probabilities, acceptance has a "yes-no" or categorical character. Acceptance may nonetheless be provisional, reexamined each time new information arrives. Abductive logic is thus relevant to the strategies a scientist can use to reach stable acceptance of a true and interesting theory.

In the present volume, we consider primarily deductive logic. Only one chapter is reserved for inductive logic, in the form of *probability theory*.[6] Indeed, our discussion of deductive logic will be narrowly focussed on its most elementary incarnation, namely, *Sentential Logic*. The latter subject analyzes the secure inferences that can be represented in an artificial language that we shall presently study in detail. Such use of an artificial language might appear strange to you inasmuch as logic is supposed to be an aid to ordinary reasoning. Since the latter typically transpires within natural language (during conversation, public debate, mental "dialogues," etc.), wouldn't it be better to tailor logic

---

[6] For more on inductive logic, see [91, 39]. For abductive logic see the internet resource:

> http://www.princeton.edu/∼osherson/IL/ILpage.htm,

and references cited there.

to English from the outset? Although such a strategy seems straightforward, it has proven awkward to implement. Before entering into the details of Sentential Logic, it will be well to discuss this preliminary point.

1.4 Artificial languages

To explain why logic has recourse to artificial rather than natural languages, let us focus on deduction (although induction or abduction would serve as well). One goal of deductive logic is to distinguish secure from insecure inferences in a systematic way. By "systematic way" is meant a method that relies on a relatively small number of principles to accurately categorize a broad class of inferences (as either secure or not). It is difficult, however, to formulate comprehensive principles of secure inference between sentences written in English (and similarly for other languages). The superficial structure of sentences seems not to reveal enough about their meanings. Here are some illustrations of the problem. In each case we exhibit a secure inference, followed by a general principle that it suggests. Then comes the counterexample to the principle. The first illustration is from Sains [88, p. 40].

Secure argument: Human beings are sensitive to pain. Harry is a human being. So, Harry is sensitive to pain.

Generalization: X's are Y. A is an X. So A is Y.

Counterexample: Human beings are evenly distributed over the earth's surface. Harry is a human being. So Harry is evenly distributed over the earth's surface.

Do you get the point? The first argument represents a secure inference. But the explanation for this fact seems not to reside in the superficial form of the sentences in the argument. (The form is shown by the schematic sentences in the middle of the display.) That the form doesn't explain the security of the original inference is revealed by the bottom argument. It has the same form, but the inference is not secure. Indeed, if "Harry" denotes a person then both premises are true yet the conclusion is false! So the inference can't be secure.

The next two examples are drawn from Katz [58, pp. xvi-xvii].

Secure argument: There is a fire in my kitchen. My kitchen is in my house. Hence, there is a fire in my house.

Generalization: X is in Y. Y is in Z. Hence, X is in Z.

Counterexample: There is a pain in my foot. My foot is in my shoe. Hence, there is a pain in my shoe.

Secure argument: Every part of the toy is silver. Hence, the toy itself is silver.

Generalization: Every part of X is Y. Hence X itself is Y.

Counterexample: Every part of the toy is little. Hence the toy itself is little.

A final example is due to Nickerson [76].

Secure argument: Professional teams are better than college teams. College teams are better than high school teams. Therefore, professional teams are better than high school teams.

Generalization: X is better than Y. Y is better than Z. Therefore, X is better than Z.

Counterexample: Nothing is better than eternal happiness. A ham sandwich is better than nothing. Therefore, a ham sandwich is better than eternal happiness.

The shifting relation between the form of a sentence and its role in inference has been appreciated by philosophers for a long time. For example, Wittgenstein (1922, p. 37) put the matter this way.

> "Language disguises thought. So much so, that from the outward form of the clothing it is impossible to infer the form of thought beneath it, because the outward form of the clothing is not designed to reveal the form of the body, but for entirely different purposes."

Similar remarks appear in Tarski [102, p. 27] and in other treatises on logic.

Yet other features of natural language make them difficult to work with. For one thing, sentences often express more than a single meaning; they are *ambiguous*. Consider the sentence "Punching kangaroos can be dangerous." Does it imply that it's unwise to punch kangaroos, or that one should stand back when a kangaroo starts punching? In this case the ambiguity is resolved by considering the lexical category of the word "punching."[7] If it is a gerund (like "writing to") then we get the first interpretation. We get the second if it is an adjective (like "colossal"). But other cases of ambiguity cannot be resolved on this basis. Consider: "Every basketball fan loves a certain player." Does it justify the conclusion that there is a single, superstar (say, M.J.) that every fan loves? Or does it just mean that for every fan there is a player the fan loves (perhaps the one that most resembles him or her). Both interpretations are in fact possible. This example is interesting because the competing interpretations don't depend on alternative lexical categorizations of any word in the sentence. Rather, something like the relative priority of the words "every" and "a" seems to determine which interpretation comes to mind. The same is true for "The professor was amazed by the profound remarks." Does this allow us to conclude that all the remarks were profound, or that it was just the profound ones that amazed the professor (the other remarks being scoffed at)?

Indeterminacy of meaning in natural language takes several forms, not just the kind of ambiguity discussed above. Consider the sentence:

> If Houston and Minneapolis were in the same state, then Houston would be a lot cooler.

Is this true, or is it rather the case that Minneapolis would be a lot warmer? Or would there be a much larger state? There are no facts about geography that settle the matter.[8] Our puzzlement seems instead to derive from some defect in the meaning expressed. Yet another kind of indeterminacy is seen in sentences

[7] The Merriam-Webster Dictionary offers the following definition of the word *lexical*: of or relating to words or the vocabulary of a language as distinguished from its grammar and construction.

[8] Adapted from Goodman [37, §1.1].

involving vague predicates, such as "Kevin Costner (the actor) is tall." It seems difficult to resolve the matter of Kevin's tallness on the basis of measurements we might make of him. No matter what we find, we may well be uncomfortable whether we declare the sentence true, or declare it false.

One reaction to the complexity of natural language is to study it harder. Indeed, much work is currently devoted to discovering linguistic principles that predict which inferences are secure, which sentences are ambiguous, which are grammatically well-formed, and so forth. Many languages have come in for extensive analysis of this sort. (See [16, 63], and references cited there.) Another reaction to the unruly character of English is to substitute a more ruly, artificial language in its place. The secure inferences are then characterized within this simpler context. Such was the path chosen by the creators of mathematical logic, starting from the middle of the nineteenth century. (For the early history of this movement, see [97].)

The most fundamental of the artificial languages studied in logic will be presented in Chapter 3 below. It will occupy us throughout the entire book. You will find the language to be neat and odorless, free of ambiguity and vagueness, and agreeable to the task of distinguishing secure from insecure inferences within it. The only perplexity that remains concerns its relation to ordinary language and the reasoning that goes on within it. For, the use of an artificial language does not resolve the complexities of inference within natural language, but merely postpones them. To get back to English, we must ascertain the correspondences that exist between inference in the two languages. This will prove a knotty affair, as you will see.

Having finished these introductory remarks, our first mission is nearly to hand. It is to present the language of Sentential Logic, central to our study of deduction and induction. But we must ask you to tolerate a brief delay. Our work will be easier if we first review some conventions and principles concerning *sets*. The next chapter is devoted to this topic. Then we'll be on our way.

(3) EXERCISE: Explain in unambiguous language the different meanings you discern in the following sentence.

"An athlete is loved by every basketball fan."

(4) EXERCISE: How many meanings do you detect in the following sentence, appearing in Elton John's *Kiss The Bride*?

> "Everything will never be the same again."

Do you think that the ambiguity was an attempt by Mr. John to be intellectually provocative, or was he just out to lunch?

(5) EXERCISE: Which of the following sentences do you find ambiguous (and what meanings are present)?

> Many arrows didn't hit the target.
>
> The target wasn't hit by many arrows.

These kinds of sentences are much discussed by linguists. (See [63, 16].)

(6) EXERCISE: Here is another example from Katz [58]. Consider the inference:

> Today I ate what I bought in the store last week. I bought a small fish at the store last week. Hence, today I ate a small fish.

What general principle does it suggest? Is there a counterexample?

(7) EXERCISE: Consider again the principle

> Every part of X is Y. Hence X itself is Y,

to which we presented a counterexample above. What further conditions can be imposed on Y to ensure the validity of the schema? Here are some possibilities. Which do the trick?[9]

(a) Y is *expansive*, that is, Y is satisfied by every whole thing any of whose parts satisfies Y (for example, "is large" or "has water inside of it").

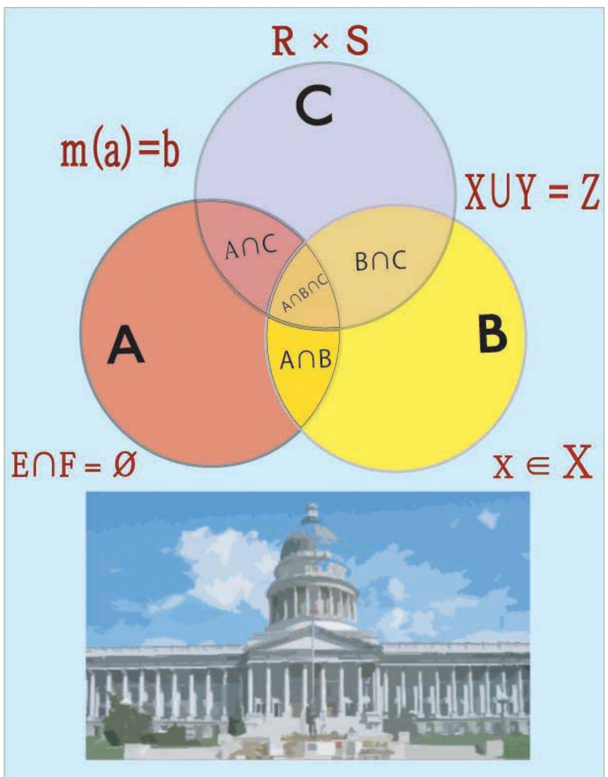
[9] The conditions come from Goodman [36, §II,4]. By an object *satisfying* Y we mean that Y is true of the object. For example, most fire engines satisfy the predicate "red."

(b) Y is *dissective*, that is, Y is satisfied by every part of a thing that satisfies Y (e.g., weighs less than the Statue of Liberty).

(c) Y is *collective*, that is, Y is satisfied by any object that can be decomposed into objects that satisfy Y (for example, "is silver," and "belongs to Bill Gates").

Chapter 2

Bits of set theory



Although it won't be to everyone's taste, we need some elementary concepts from set theory in order discuss Sentential Logic. If you find that you can't keep your eyes focussed on what follows then skip it; go directly to Chapter 3. When we invoke an idea about sets for the first time, we'll give you a citation to the relevant section in this chapter.

We hope that you've already seen the material about to be reviewed. The present text is not the best place to study it for the first time.[1] What follows is self-contained, but nonetheless in the spirit of memory-revival.

2.1 Sets and elements

At the risk of getting off on the wrong foot, we admit to being unable to define "set." The concept is just too basic. Instead of a definition, mathematicians typically say that sets are "definite collections" of objects. The objects that comprise a set are known as its *members* or *elements*. Thus, the set S of United States Senators in 2004 is the collection of people consisting of Diane Feinstein, John Kerry, Hillary Clinton, and so forth. As you know, S has 100 members. We write "Joseph Bieden $\in S$" to indicate that Joseph Biden is an element of S, and "Jacques Chirac $\notin S$" to deny that Jacques Chirac is a member. Braces are used to indicate the members of a set. Thus, $\{a, e, i, o, u\}$ denotes the set of vowels in the alphabet. A set with just one member is called a *singleton* set. For example, $\{TonyBlair\}$ is a singleton set, containing just the famous British Prime Minister.[2]

It is crucial to note that sets are not ordered. Thus, $\{a, e, i, o, u\}$ is *identical* to the set denoted by $\{u, o, i, e, a\}$. All that matters to the specification of a set is its members, not the order in which they happen to be mentioned. *Sets with the same members are the same set.*

In the chapters to follow, our use of sets will always be accompanied by a clear choice of *domain* or *universal set*. By a "domain" is meant the set of all

[1] Introductions include [68, 84].

[2] In this example we denoted Tony Blair with two words, "Tony" and "Blair." But the set has just one member (the doughty Mr. Blair). We use commas to separate members in a set that is listed between braces. Thus, $\{\text{Tony Blair}, \text{Jacques Chirac}, \text{Geronimo}\}$ has three members.

elements that are liable to show up in any set mentioned in the surrounding discussion. For example, in discussing electoral politics in the United States, our domain might be the set of U. S. citizens. In discussing spelling, our domain might be the set of all English words (or perhaps the set of all finite strings of letters, whether they are words or not). In this chapter, we'll sometimes leave the domain implicit. When we need to remind ourselves that all elements are drawn from a fixed universal set, we'll denote the universal set by $\mathcal{U}$.

To carve a set out of $\mathcal{U}$ we sometimes use "set-builder notation." It relies on the bar $\mid$ to represent a phrase like "such that" or "with the property that." To illustrate, suppose that our domain is the set of natural numbers $\{0, 1, 2, \ldots\}$, denoted by N.[3] Then the expression $\{x \in N \mid x \geq 10\}$ denotes the set of natural numbers that are greater than or equal to 10. You can read the thing this way: "the set of elements (say x) of N with the additional property that the element (namely, x) is greater than or equal to 10" — or as "the set of x in N such that $x \geq 10$." For another example, if E is the set of even members of N then $\{x \in E \mid x \text{ is divisible by } 5\}$ denotes the set $\{10, 20, 30, \ldots\}$. Because set-builder notation is so important, we'll illustrate it one more time. If $\mathcal{U}$ is the set of NBA players, then $\{x \mid x \text{ earned more than a million dollars in 2004}\}$ would be the set that includes Allen Iverson and Jason Kidd, among others. If we want to cut down the number of players under discussion, we could write $\{x \in \text{New Jersey Nets} \mid x \text{ earned more than a million dollars in 2004}\}$ to denote just the members of the Nets who earned more than a million dollars in 2004 (Jason Kidd and some others, but not Allen Iverson). The latter set could also be written:

$$\{x \mid x \text{ plays for the Nets and earned more than a million dollars in 2004}\}.$$

Style and convenience determine how a given set is described.

2.2 Subsets

If sets A and B have the same members, we write $A = B$; in this case, A and B are the same set. For example, let A be the set of former U.S. presidents alive

[3] The symbol "$\ldots$" means: "and so forth, in the obvious way."

in 2003, and let B be $\{\text{Ford}, \text{Carter}, \text{Reagan}, \text{G. H. Bush}, \text{Clinton}\}$. Then $A = B$.

Given two sets A, B and a domain $\mathcal{U}$, we write $A \subseteq B$, and say that "A is a subset of B," just in case $\mathcal{U}$ has no elements that belong to A but not B. Let's repeat this officially.

(1) DEFINITION: Let domain $\mathcal{U}$ and sets A, B be given. Then $A \subseteq B$ just in case there is no $x \in \mathcal{U}$ such that $x \in A$ but $x \notin B$. In this case, A is said to be a *subset* of B.

For example, $\{3, 4, 6\} \subseteq \{3, 4, 5, 6\}$. Also, if A is the set of vowels and B is the entire alphabet then $A \subseteq B$. It should be clear that:

(2) FACT: For any sets A, B, $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.

Definition (1) implies that every set is a subset of itself. That is, $B \subseteq B$. This is because there is no member of B that fails to be in B. B is called the *improper* subset of B. A *proper* subset of B is any of its subsets that is not improper, that is, any of B's subsets save for B itself. Officially:

(3) DEFINITION: Let sets A, B be given. Then $A \subset B$ just in case $A \subseteq B$ and $A \neq B$. In this case, A is said to be a *proper* subset of B.

For example, $\{2, 4\} \subset \{2, 3, 4\}$. Notice that we use $A \neq B$ to deny that $A = B$; in general, a stroke through a symbol is used to assert its denial. An equivalent formulation of Definition (3) is that $A \subset B$ just in case $A \subseteq B$ and $B \not\subseteq A$. The following fact should also be evident.

(4) FACT: For any sets A, B, $A \subset B$ if and only if $A \subseteq B$ and there is $x \in B$ such that $x \notin A$.

In the example $\{2, 4\} \subset \{2, 3, 4\}$, the number 3 is the x in Fact (4).

Notice the difference between the symbols $\subseteq$ and $\subset$. The bottom stroke in $\subseteq$ suggests that equality is left open as a possibility; that is, $A \subseteq B$ is compatible with $A = B$ whereas $A \subset B$ excludes $A = B$. Of course, $A = B$ implies $A \subseteq B$.

For sets A, B, $A \supseteq B$ means $B \subseteq A$. The symbol $\subseteq$ is turned around to make $\supseteq$. If $A \supseteq B$ then we say that A is a *superset* of B. Likewise, we may write $A \supset B$ in place of $B \subset A$, and say that A is a *proper superset* of B.

(5) EXERCISE: Suppose that $A = \{a, b, c, d\}$. List every proper subset of A.

2.3 Complementation

The next few sections provide ways of defining new sets from given sets.

(6) DEFINITION: Given sets A, B, we let $A - B$ denote the set of $x \in A$ such that $x \notin B$.

For example, if $A = \{3, 5, 8\}$ and $B = \{5, 9, 10, 15\}$ then $A - B = \{3, 8\}$ and $B - A = \{9, 10, 15\}$. Note that typically, $A - B \neq B - A$. In Definition (6), A might be the whole domain $\mathcal{U}$. This case merits special treatment.

(7) DEFINITION: Given set B, we let $\overline{B}$ denote $\mathcal{U} - B$, namely, the set of $x \in \mathcal{U}$ such that $x \notin B$. The set $\overline{B}$ is called the *complement* of B.

For example, if $\mathcal{U}$ is the set of letters, then the complement of the vowels is the set of consonants. If $\mathcal{U}$ is the set of natural numbers $0, 1, 2, 3 \ldots$ then $\overline{\{0, 1, 3, 4, 5\}}$ is the set of natural numbers greater than 5.

2.4 Intersection

$A \cap B$ denotes the set of elements common to A and B. Officially:

(8) DEFINITION: Let domain $\mathcal{U}$ and sets A, B be given. We let $A \cap B$ denote the set of $x \in \mathcal{U}$ such that $x \in A$ and $x \in B$. The set $A \cap B$ is called the *intersection* of A and B.

For example, $\{5, 8, 9\} \cap \{2, 8, 9, 11\} = \{8, 9\}$. Also, $\{2, 8, 9, 11\} \cap \{5, 8, 9\} = \{8, 9\}$, and more generally $A \cap B = B \cap A$ is always the case.

2.5 Union

$A \cup B$ denotes the set that results from pooling the members of A and B. That is, $A \cup B$ is the set whose members appear in at least one of A and B. Officially:

(9) DEFINITION: Let domain $\mathcal{U}$ and sets A, B be given. We let $A \cup B$ denote the set of $x \in \mathcal{U}$ such that $x \in A$ or $x \in B$ (or both). The set $A \cup B$ is called the *union* of A and B.

For example, $\{4,2,1\} \cup \{3,4,6\} = \{4,2,1,3,6\}$. Don't be tempted to write $\{4,2,1\} \cup \{3,4,6\} = \{4,4,2,1,3,6\}$ thereby signaling the occurrence of 4 in both sets of the union. Sets are determined by their members, so $\{4,4,2,1,3,6\}$ is just the set $\{4,2,1,3,6\}$ since they have the same members. It's confusing to use $\{4,4,2,1,3,6\}$ in place of $\{4,2,1,3,6\}$ since it invites the mistaken idea (not intended in the present example) that the two sets $\{4,2,1\}$, $\{3,4,6\}$ contain *different copies* of the number 4. Of course, we always have $A \cup B = B \cup A$.

(10) EXERCISE: In the domain of numbers from 1 to 10, let $A = \{2,4,6,9\}$, $B = \{3,4,5,6,1\}$, $C = \{8,10,1\}$. What are the sets $A \cup B$, $A \cap B$, $A \cap C$, $B - A$, $A - B$, $\overline{A \cup C}$, and $\overline{B}$?

2.6 The empty set

Here is an important postulate about sets that we adopt without further discussion.

(11) POSTULATE: There is a set without any members.

The postulate immediately implies that there is exactly one set without any members. For suppose that A and B are both sets without any members. Then A and B have the same members, namely, none. Consequently, $A = B$ since sets with the same members are identical. In other words, A and B are identical; they are the same set. So, there can't be more than one set without any members. We give this unique set a name and a symbol.

(12) DEFINITION: The set without any members is called the *empty set* and denoted $\emptyset$.

We have a notable consequence of Definition (1).

(13) FACT: For every set A, $\emptyset \subseteq A$.

The fact follows from the absence of members of $\emptyset$, which implies that no members of $\emptyset$ fail to be members of A. This is all we need to infer that $\emptyset \subseteq A$ [see Definition (1)].

Of course, for every nonempty set B, $\emptyset \subset B$. That is, the empty set is a proper subset of every other set. And $\emptyset \subseteq \emptyset$.

When the intersection of sets is empty, we call them "disjoint." That is:

(14) DEFINITION: Suppose that sets A and B are such that $A \cap B = \emptyset$. Then we say that A and B are *disjoint*.

For example, the set of even integers and the set of odd integers are disjoint.

(15) EXERCISE: Mark the following statements as true or false.

(a) For all sets A, $A \cap \emptyset = A$.
(b) For all sets A, $A \cup \emptyset = A$.
(c) For all sets A, $A \cap \emptyset = \emptyset$.
(d) For all sets A, $A \cup \emptyset = \emptyset$.
(e) $\overline{\mathcal{U}} = \emptyset$.
(f) $\overline{\emptyset} = \mathcal{U}$
(g) For all sets A, $A \cup \mathcal{U} = \mathcal{U}$.
(h) For all sets A, $A \cap \mathcal{U} = \mathcal{U}$.
(i) For all sets A, $A \cup \mathcal{U} = A$.
(j) For all sets A, $A \cap \mathcal{U} = A$.

(16) EXERCISE: Which of the following statements are true, and which are false? (The expression "if and only if" can be read "just in case" or "exactly when.") Each statement is a claim about all sets A, B, C.

(a) $A \subseteq B$ iff $A \cap \overline{B} = \emptyset$.

(b) $A \subseteq B$ iff $\overline{A} \cup B = \mathcal{U}$.

(c) $A \subseteq B$ iff $B \subseteq A$.

(d) $A \subseteq B$ iff $\overline{B} \subseteq \overline{A}$.

(e) $A \subseteq B$ iff $A - B = \emptyset$.

(f) $A \subseteq B$ iff $B - A = \emptyset$.

(g) $\overline{A \cup B} = \overline{A} \cup \overline{B}$

(h) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(i) $(A \cup B) \cap (A \cup \overline{B}) = A$

(j) $A \cup (A \cap B) = A$

(k) $A \cap (A \cup B) = A$

(l) $A \cup (B - C) = (A \cup B) - (A \cup C)$

(m) $A \cap (B - C) = (A \cap B) - (A \cap C)$

2.7 Power sets

Let $S = \{a, b, c, \}$. Keeping in mind both $\emptyset$ and S itself, we may list all the subsets of S as follows.

(17)
$$\begin{array}{cccc} \emptyset & \{a\} & \{b\} & \{c\} \\ \{a,b\} & \{a,c\} & \{b,c\} & \{a,b,c\} \end{array}$$

We call the set consisting of the sets in (17) the "power set" of S. The power set of a set is thus composed of sets; its members are themselves sets. Officially:

(18) DEFINITION: Given a set S, the *power set* of S is the set of all subsets of S. It is denoted by *pow*(S).

Thus *pow*(S) equals the set of sets listed in (17), in other words:

$$\{\ \emptyset, \{a\}, \{b\}, \{c\}, \{a,b\}, \{a,c\}, \{b,c\}, \{a,b,c\}\ \}.$$

Why are there eight members of *pow*(S)? Well, each member of S can be either in or out of a given subset. There are three such binary decisions, and they are made independently of each other. That yields $2 \times 2 \times 2$ or 2^3 possibilities (and $2^3 = 8$). For example, deciding "out" in all three cases yields $\emptyset$; deciding "in" all three times yields S itself. More generally:

(19) FACT: There are 2^n subsets of a set with n members. In other words, if S has n members then *pow*(S) has 2^n members.

By the way, the terms "collection" and "family" are sometimes used in place of "set," especially to denote sets whose members are also sets or other complicated things. So, we could have formulated Definition (18) more elegantly, as follows.

Given a set S, the *power set* of S is the collection of all subsets of S.

(20) EXERCISE:

(a) What is *pow*$(2, \{2,3\})$?

(b) What is *pow*$(\emptyset)$?

2.8 Partitions

A "partition" of a set A is breakdown of A into pieces that don't overlap and include every member of A; moreover, none of the pieces is allowed to be empty. For example, one partition of the 10 digits is $\{\ \{0,2,4,6,8\}, \{1,3,5,7,9\}\ \}$. The latter object is a set with two members, each of which happens to be a set; in fact, a partition of A is a set of subsets of A. Another partition of the digits is $\{\ \{0,2,6,8\}, \{1,3,5,7\}, \{4,9\}\ \}$; still another is

$$\{\ \{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}, \{7\}, \{8\}, \{9\}\ \},$$

made up of singletons (this is the "finest" partition of the digits). Yet another is $\{\ \{0,1,2,3,4,5,6,7,8,9\}\ \}$ (the "grossest" partition). With these examples in mind, you should be able to understand the official definition of "partition."

(21) DEFINITION: Let X be a nonempty set, and let $\mathcal{P}$ be a collection of sets. Then $\mathcal{P}$ is called a *partition* of X just in case the following conditions are met.

(a) Each set in $\mathcal{P}$ is a nonempty subset of X. That is, for all $Y \in \mathcal{P}$, $\emptyset \neq Y$ and $Y \subseteq X$.

(b) Every pair of sets in $\mathcal{P}$ is disjoint. That is, for all distinct $Y, Z \in \mathcal{P}$, $Y \cap Z = \emptyset$.

(c) Every member of X falls into some member of $\mathcal{P}$. That is, for all $x \in X$ there is $Y \in \mathcal{P}$ with $x \in Y$. (Hence, the union of the sets in $\mathcal{P}$ equals X).

We also say that $\mathcal{P}$ *partitions* X. The members of $\mathcal{P}$ are called the *equivalence classes* of the partition. (Thus, equivalence classes are subsets of X.)

For another example, let X be the set of U. S. citizens with permanent residence in a state of the United States. Let TEXAS be the set of all U. S. citizens with permanent residence in Texas, and likewise for the other states. Let $\mathcal{P}$ be the collection of all these fifty sets of people, one for each state of the union. Then you can verify that $\mathcal{P}$ is a partition of X. The equivalence classes are the sets TEXAS, IDAHO, etc.

Again, the set of living United States citizens (no matter where they live) is partitioned into age-cohorts according to year of birth. One equivalence class contains just the people born in 1980, another contains just the (few) people born in 1900, etc. There is no equivalence class corresponding to the people born in 1800 since equivalence classes cannot be empty [by (21)a].

(22) EXERCISE: Let $A = \{a, e, i, o, u\}$. Which of the following collections of sets are partitions of A?

(a) $\{ \{a, e, i\}, \{i, o, u\} \}$
(b) $\{ \{a, e, i\}, \{o, u\}, \emptyset \}$
(c) $\{ \{a, e\}, \{i\}, \{u\} \}$
(d) $\{ \{a, e\}, \{i\}, \{u, o\} \}$

2.9 Ordered pairs and Cartesian products

We said in Section 2.1 that sets are not ordered, for example, $\{2, 9\} = \{9, 2\}$. Because order is often critical, we introduce the idea of an *ordered pair*, $\langle y, z\rangle$. Two ordered pairs $\langle y, z\rangle$ and $\langle w, x\rangle$ are identical just in case $y = w$ and $z = x$.

We could get fancy here, and define the idea of "ordered pair" from our basic idea of a set. But ordered pairs seem clear enough to let them stand on their own. The sequence consisting of the Yankees and the Dodgers, in that order, is the ordered pair $\langle \text{Yankees}, \text{Dodgers}\rangle$. No other pair $\langle x, y\rangle$ is the same as this one unless x is the Yankees and y is the Dodgers. Notice the use of parentheses instead of braces when we denote ordered pairs (instead of mere sets). Also, notice that we sometimes drop the qualifier "ordered" from the expression "ordered pair."

Given a pair $\langle x, y\rangle$, we say that x occupies the *first coordinate*, and y the *second coordinate*. For example, Hillary and Bill are in the first and second coordinates of $\langle \text{Hillary}, \text{Bill}\rangle$.

Suppose $A = \{a, b\}$ and $B = \{1, 2, 3\}$. How many ordered pairs can be made from these two sets, supposing that a member of A occupies the first coordinate and a member of B occupies the second? The answer is 6. Here are all such pairs, drawn up into a set of 6 elements: $\{ \langle a, 1\rangle, \langle a, 2\rangle, \langle a, 3\rangle, \langle b, 1\rangle, \langle b, 2\rangle, \langle b, 3\rangle \}$. More generally:

(23) DEFINITION: Let sets A and B be given. The set of all ordered pairs $\langle x, y\rangle$ where $x \in A$ and $y \in B$ is denoted $A \times B$ and called the *Cartesian product* of A and B.

If M is the set of men on Earth and W is the set of women then $M \times W$ is the set of all pairs consisting of a man followed by a woman. $W \times M$ is the set of all pairs consisting of a woman followed by a man. You can see that $M \times W \neq W \times M$. Indeed, these two sets are disjoint (assuming M and W are disjoint).

(24) EXERCISE:

(a) Let $A = \{a, b, c\}$ and $B = \{a, b, d\}$, how many members are there in $A \times B$?

(b) What is $A \times \emptyset$?

2.10 Mappings and functions

Intuitively, a "mapping" from a set A to a set B is a way of associating with each member of A a single member of B. For example, we might map the set of NBA basketball players to the salaries each is currently scheduled to receive. This mapping consists of pairs $\langle x, y \rangle$ consisting of a player x and a number y such that x earned y dollars last year. The pair $\langle$Allen Iverson, $\$10,000,000\rangle$ is one such pair (we're just guessing). More precisely:

(25) DEFINITION: By a *mapping* M from a set A to a set B is meant a subset of $A \times B$ with the property that for all $x \in A$ there is exactly one pair in M that has x in the first coordinate. We write $M : A \rightarrow B$ to express the fact that M is a mapping from A to B. Given $a \in A$, we write $M(a)$ to denote the member of B that M pairs with a.

Pursuing our illustration, let A be the set of NBA players and let B be the set of potential salaries (positive numbers). Then the map $M : A \rightarrow B$ includes the pairs $\langle$Allen Iverson, $\$10$ million$\rangle$, $\langle$Shaq O'Neal, $\$50$ million$\rangle$, etc. We may thus write:

$$M(\text{Allen Iverson}) = \$10 \text{ million}, \quad M(\text{Shaq O'Neal}) = \$50 \text{ million}, \text{ etc.}$$

Iverson, of course, only gets one salary (he doesn't get paid both 10 million and also 9 million dollars as his NBA salary). So we don't expect to see any other pair in M that starts with Iverson, different from the pair $\langle$Allen Iverson, $\$10$ million$\rangle$. That is what's meant in Definition (25) by saying that for each $x \in A$ there is *exactly one pair* in M that has x in first coordinate. In other words, $\$10$ million is Iverson's *unique* NBA salary. It is for this reason that we are allowed to write $M(\text{Iverson}) = \$10$ million.

Imagine now that Karl Malone is in negotiation mode and doesn't yet have a salary. Then we can associate no second coordinate to a pair that starts with Malone. In this case, we drop one pair from our mapping and say that it is "undefined" on Malone. But to do this, we must give up the "mapping" terminology, and talk instead of a "function" from the NBA players to their salaries. Functions (unlike mappings) are allowed to be missing some pairs. Officially:

(26) DEFINITION: By a *function* F from a set A to a set B is meant a subset of a mapping from A to B. If $\langle x, y\rangle \in F$ then we say that F is *defined* on x and write $F(x) = y$. If there is no pair with x in first coordinate then we say that F is *undefined* on x.

In our story, $F(\text{Iverson})$ is defined, and the value of F on Iverson is $\$10$ million. But $F(\text{Malone})$ is undefined.

Notice that mappings are special kinds of functions since F may be the improper subset of M in Definition (26); see Section 2.2.

(27) EXERCISE: Suppose that A has n members and that B has m members.

(a) How many mappings are there from A to B ?

(b) How many functions are there from A to B ?

Of course, your answer must be stated in terms of n and m (since no actual numbers are supplied).

2.11 Mathematical induction

By *natural numbers*, we mean the set $\{0, 1, 2, \ldots\}$. As a final topic in this chapter, we consider a method for proving that every natural number has some property. For example, let P be the property of a number n that holds just in case all the numbers less than or equal to n sum up to $n(n+1)/2$. To affirm that P is true of a number n, we write $P(n)$. Thus in our example:

(28) $$P(n) \text{ if and only if } 0+1+2+\ldots+n = \frac{n(n+1)}{2}.$$

In Chapter 7 we rely on a principle that can be formulated this way:

(29) MATHEMATICAL INDUCTION: For any property P, to prove that $P(n)$ for all $n \in N$, it suffices to demonstrate the following things.

(a) BASIS STEP: $P(0)$

(b) INDUCTION STEP: for all $k \in N$, if $P(k)$ then $P(k+1)$.

The idea is that proving both parts of (29) allows us to "walk up" the natural numbers. The first step is guaranteed by (29)a. Then every further step is guaranteed by an application of (29)b. It goes like this:

$P(0)$ is true by the Basis Step.
Since $P(0)$, $P(1)$ is true by the Induction Step.
Since $P(1)$, $P(2)$ is ensured by another use of the Induction Step.
Since $P(2)$, $P(3)$ is ensured by another use of the Induction Step.
. . . and so forth.

Let us illustrate how to use mathematical induction by letting P be defined as in (28). We want to show that $P(n)$ for all $n \in N$. The Basis Step requires verifying that $P(0)$, in other words: $0 = 0(1)/2$, which is evident. For the Induction Step, we assume that $P(k)$ is true for some arbitrary $k \in N$. This assumption is called the *induction hypothesis*. Then we exploit the induction hypothesis to demonstrate $P(k+1)$. In our case, we assume:

Induction Hypothesis: $0+1+2+\ldots+k = \frac{k(k+1)}{2}$,

and we must demonstrate:

$$0+1+2+\ldots+k+(k+1) = \frac{(k+1)(k+2)}{2}.$$

To prove $P(k+1)$, we calculate:

$$0 + 1 + 2 + \ldots + k + (k+1) = \frac{k(k+1)}{2} + (k+1)$$

$$= \frac{k(k+1) + 2(k+1)}{2} = \frac{(k+1)(k+2)}{2}.$$

The first equality is obtained from the Induction Hypothesis. The next two are just algebra.

(30) EXERCISE: Use mathematical induction to prove the following claims.

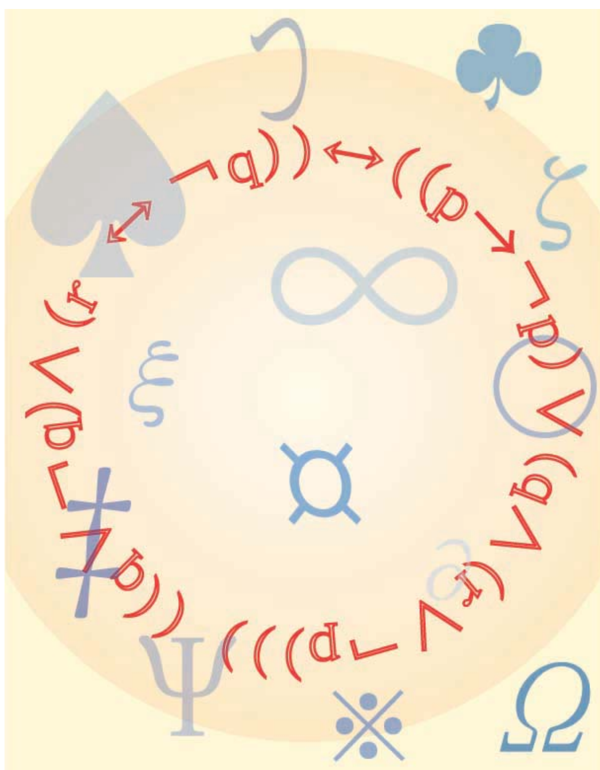
(a) For all $n \in N$, $2^n > n$.

(b) For all sets S, if S has n members then *pow*(S) has 2^n members.

At last we are done with these maddening preliminaries! Before some new distraction arises, let us head straight for the language of logic.

Chapter 3

The language of Sentential Logic



3.1 The syntax project

In Chapter 1 we promised you an artificial language, and now we are going to deliver. It will be called *the language of sentential logic,* or just $\mathcal{L}$ for short. When we are done, you'll see that $\mathcal{L}$ is nothing but a set of objects called *formulas*. Once we explain what a formula is, you will know what the language of sentential logic is (namely, it is the set of all formulas).

Intuitively, formulas are the "sentences" of $\mathcal{L}$, They are supposed to be analogous to the sentences of *natural* languages like English and Chinese. Giving substance to this intuition requires explaining what an English sentence is. Dictionary definitions are too vague to provide the explanation we are looking for. For example, we find the following characterization of *sentence* in the dictionary [1].

> A grammatical unit that is syntactically independent and has a subject that is expressed or, as in imperative sentences, understood and a predicate that contains at least one finite verb.

Instead of a one-line definition, what's needed is a detailed theory about the strings of words that make up English sentences. Such a theory would help to specify the *syntax* of the language.[1] If the theory were detailed and explicit, it would provide a systematic means of distinguishing between the strings of English words that form grammatical sentences and those that don't. In the first category are strings like

(1) (a) John was easy to please.

(b) John was easy to like.

(c) To please John was easy.

(d) John was eager to please.

[1]The definition of "syntax" according to the same dictionary [1] is somewhat more useful, namely: "The study of the rules whereby words or other elements of sentence structure are combined to form grammatical sentences."

whereas the second class includes:

(2) (a) *John was eager to like.

(b) *To please John was eager.[2]

The formulas of $\mathcal{L}$ are analogous to grammatical sentences of English, like (1). Strings of symbols that do not make formulas are analogous to ungrammatical strings of English words, as in (2).

Providing an illuminating account of grammaticality in English turns out to be a knotty affair. The examples in (1) and (2) suggest that more is at stake than having the right sequence of parts-of-speech since (1)b and (2)a have the same sequence but only one is grammatical [and similarly for (1)c versus (2)b]. Indeed, grammaticality in English seems to get more complicated the more you think about it. Consider the strings in (3), below, involving the word "that." A satisfactory account of English syntax would need to explain why "that" can be suppressed in (3)a,b without loss of grammaticality, but not in (3)c,d — and also why it *must* be suppressed in (3)e.[3]

(3) (a) Irving believed that pigs can fly.

(b) The pain that I feel is most unpleasant.

(c) That sugar is sweet is obvious to everyone.

(d) The dog that bit me is missing now.

(e) *The dog Mary feared that bit me is missing now.

In fact, no one has yet provided a systematic way of predicting which strings are grammatical in English.[4] This is one reason for introducing the artifical language $\mathcal{L}$, whose syntax is designed to be much simpler.

[2]These examples are drawn from the famous discussion in Chomsky [19]. An asterisk before a sentence is the usual linguist's convention for denoting an ungrammatical string of words.

[3]These examples come from Weisler & Milekic [106, p. 130]. While you're at it, ask yourself why *John ran up a big hill*, *John ran up a big bill*, *Up a big hill John ran* are all grammatical English whereas *Up a big bill John ran* is not.

[4]For an introduction to the study of syntax in natural languages like English, see [64, 32].

In the foregoing discussion, the formulas of $\mathcal{L}$ have been placed in analogy with the grammatical sentences of English. The analogy can be sharpened by recognizing that many sentences do not make assertions but rather ask questions, express puzzlement, issue commands, and so forth. Thus,

(4) Hillary will be reelected to the Senate in 2006.

makes an assertion, whereas this is not the case for any of the following sentences.

(5) (a) Will Hillary be reelected to the Senate in 2006?
(b) Hillary will be elected to the Senate in 2006!? (I thought the next election was in 2008.)
(c) Hillary, please get yourself elected to the Senate in 2006!

Sentences like (4) that make assertions are called *declarative* in contrast to the *interrogative* and *imperative* sentences in (5). The formulas of $\mathcal{L}$ are best thought of as artificial counterparts to just the declarative sentences of English. This is quite a restriction. It's still not enough for our purposes, however, since not even all of the declarative sentences have analogues in $\mathcal{L}$. For, among the declarative sentences of English are many that have no apparent meaning, and hence seem to be neither true nor false. One example (from Chomsky [18]) is:

(6) Colorless green ideas sleep furiously.

Although grammatical and declarative, (6) does not provide raw material for the kind of reasoning that we wish to analyze using the tools of logic. So, formulas in $\mathcal{L}$ can best be understood as corresponding to just the declarative sentences of English with clear meaning and determinate truth-value. But now a host of questions rise up before us.

- Does "determinate truth-value" mean *entirely true* or *entirely false*, or is it permitted that the sentence have some intermediate degree of truth?

- In the latter case, what does it mean for a sentence to be partly true?
- Is (6) really grammatical, or does its lack of sense put it in the same category as (2)a,b?
- Is it so clear that (6) has no truth value? It would seem that the sentence is true just in case colorless green ideas really do sleep furiously.
- Does the sentence "I am happy" have a determinate truth-value? Perhaps its truth varies with time, or with speakers.

Such questions lead into precisely the morass that the artificial language of sentential logic is designed to avoid. Lacking the courage to confront these issues head on, we'll just assume (for now) that the idea of a *declarative English sentence with determinate truth value* is clear. In particular, we assume that such sentences are either (entirely) true or (entirely) false. Within $\mathcal{L}$ everything will then be crystal clear, and we can build our logic around it. Retreating like this from English into $\mathcal{L}$ will prove illuminating but it won't protect us indefinitely from hard questions about truth and meaning. It will still be necessary to ask: What does sentential logic have to do with thought, and with the natural language often used to express thought? Addressing these questions will lead us to a plurality of truth-values, new kinds of meanings, and other exotica.

Each thing in its own time, however. For the moment, we are playing by the rules of logic, and this means we must start by defining the formulas of $\mathcal{L}$.

3.2 Vocabulary

Just as English sentences are strings of words, formulas of $\mathcal{L}$ are strings of symbols. The symbols are the *vocabulary* of $\mathcal{L}$, and we need to introduce them first. The vocabulary of $\mathcal{L}$ falls into three categories. The first consists of a finite set of symbols that stand for declarative sentences with determinate truth value (in the sense just discussed in Section 3.1). These symbols will be called "sentential variables." How many sentential variables are there? Well, it's your choice. If you want to use sentential logic to reason about many different sentences, then choose a large number of variables; otherwise, a small number suffices (but it must be greater than zero). Go ahead and decide. How many sentential variables do you want?

Excellent choice! We think we know the number you chose, but just in case we're mistaken, let us agree to use the symbol n to denote it. We henceforth assume that there are exactly n sentential variables. With this matter out of the way, the vocabulary of $\mathcal{L}$ can be described as follows.

The first category of symbols is a set $v_1, v_2 \cdots v_n$ of n *sentential variables*. In other words, for each positive $i \leq n$, v_i is a sentential variable. We usually abbreviate "sentential variable" to just "variable."

The second category consists of the five symbols $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$. Collectively, these symbols are called "sentential connectives," which we will usually abbreviate to just "connectives."

The third and final category of symbols in $\mathcal{L}$ consists of just the left parenthesis and the right parenthesis, $($, $)$.

The symbols $\wedge$, $\vee$, $\rightarrow$ and $\leftrightarrow$ are called "connectives" because they serve to connect different formulas, making a new one where two stood before. How this happens will be specified in Definition (8) below. These symbols are analogous to words like "and," "but," "although," and many others in English that serve to bind sentences to each other.[5] The symbol $\neg$ is also considered to be a connective but this is mainly by courtesy. As you'll see in Definition (8), $\neg$ applies to just one formula at a time (hence doesn't connect two of them). An analogous grammatical operation in English is appending the expression "It is not the case that" in front of a sentence.

On another matter, don't be confused by variables like v_{412} with big indexes. (The "index" is the subscript, in this case 412.) Variable v_{412} counts as a single symbol from the point of view of $\mathcal{L}$. We don't think of v_{412} as a combination of the four symbols v, 4, 1, 2 but rather as the indivisible four hundred and twelfth variable of our language.

[5] As in: "The hot dog vendor offered to make him one with everything *but* the Buddhist monk said it wasn't necessary."

In what follows we'll assume that n is greater than the subscript of any variable figuring in our discussion. (For example, if we use the variable v_5 then it is assumed that $n \geq 5$.)

3.3 Formulas

There are many ways of arranging our vocabulary into a sequence of symbols, for example:

(7) (a) $v_3 v_3 \rightarrow)$

(b) $\wedge \vee (v_{44}$

(c) $((v_{32} \wedge v_0) \rightarrow v_3)$

(d) $(v_3 \wedge v_2 \wedge \leftrightarrow v_2)$

Only some sequences are entitled to be called "members of $\mathcal{L}$," that is, "formulas." One such sequence is (7)c. It is a genuine formula whereas the other sequences in (7) are not. In this section we say precisely which sequences of symbols belong to our language $\mathcal{L}$. [Strings like (7)a may be considered "ungrammatical" on analogy with the strings in (2).]

The following definition specifies the members of $\mathcal{L}$ (i.e., the "grammatical" strings of symbols). The definition is said to be "recursive" in character since it labels some sequences as formulas in virtue of their relations to certain shorter sequences which are formulas. In the definition, we need to refer to formulas that have already been defined without specifying which they are. For example, we'll need to say things like "For every formula ..." and then say something about the formula in question. To give temporary names to formulas in such contexts we rely on letters from the Greek alphabet, notably: φ, ψ, and χ. They are pronounced "figh," "sigh," and "kigh," respectively. Other greek letters that show up for other purposes are α, β, and γ ("alpha," "beta" and "gamma"). We could have been less fancy in our notation, but couldn't resist the opportunity to complete your classical education. Here's the definition.

(8) DEFINITION:

(a) Any sentential variable by itself is a *formula*. (For example, v_{201} is a formula.)

(b) Suppose that φ is a formula. Then so is $\neg\varphi$. (For example, $\neg v_{201}$ is a formula since by the first clause of the present definition, v_{201} is a formula. Likewise, $\neg\neg v_{201}$ is a formula since $\neg v_{201}$ is a formula.)

(c) Suppose that φ and ψ are formulas. Then so is $(\varphi \wedge \psi)$. [For example, $(\neg\neg v_{201} \wedge v_{39})$ is a formula.]

(d) Suppose that φ and ψ are formulas. Then so is $(\varphi \vee \psi)$. [For example, $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ is a formula since $(\neg\neg v_{201} \wedge v_{39})$ is a formula by the preceding clause, and $\neg v_1$ is a formula by clauses (a) and (b).]

(e) Suppose that φ and ψ are formulas. Then so is $(\varphi \rightarrow \psi)$. [For example, $(\neg v_1 \rightarrow (\neg\neg v_{201} \wedge v_{39}))$ is a formula.]

(f) Suppose that φ and ψ are formulas. Then so is $(\varphi \leftrightarrow \psi)$. [For example, $(\neg v_1 \leftrightarrow (\neg\neg v_{201} \wedge v_{39}))$ is a formula.]

(g) Nothing else is a formula, just what is declared to be a formula by the preceding clauses.

Parentheses matter. Thus $(v_2 \wedge v_1)$ is a formula by (8)(a) and (8)(c). But $v_2 \wedge v_1$ is not a formula since it is missing the parentheses stipulated in the definition. Similarly, $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ is a formula but $(\neg\neg v_{201} \wedge v_{39} \vee \neg v_1)$ is not.

You see how the rules in Definition (8) allow us to build a complex formula from its parts. For example, the formula $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ is built up as follows.

(9) First, v_{201}, v_{39} and v_1 are formulas by dint of (8)a. Hence, $\neg v_1$, $\neg v_{201}$ and $\neg\neg v_{201}$ are formulas thanks to (8)b. Hence, $(\neg\neg v_{201} \wedge v_{39})$ is a formula via (8)c. Hence $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ is a formula because of (8)d.

Here is another example. We build $\neg((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$ as follows.

(10) v_3 and v_5 are formulas by (8)a. So $(v_3 \wedge v_5)$ is a formula by (8)c. v_6 and v_7 are formulas by (8)a. So $(v_6 \vee v_7)$ is a formula by (8)d. Since $(v_3 \wedge v_5)$ and $(v_6 \vee v_7)$ are both formulas, so is $((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$ by (8)f. By (8)b, $\neg((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$ is a formula.

3.4 Subformulas

Did you notice that in the course of building $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ in (9), we wrote smaller strings of symbols that are themselves formulas? For example, we wrote v_{201} and $\neg v_{201}$, which are themselves formulas. Similarly, in building $\neg((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$ in (10) we wrote the smaller formula $(v_6 \vee v_7)$. These formulas are called "subformulas" of the formula from which they are drawn. Their official definition is as follows.

(11) DEFINITION: Let φ be a formula. Any consecutive sequence of symbols in φ that is itself a formula is called a *subformula* of φ.

To illustrate, here is a list of the eight subformulas of $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$.

(12)

v_{201}	$\neg v_{201}$	$\neg\neg v_{201}$	v_{39}
$(\neg\neg v_{201} \wedge v_{39})$	v_1	$\neg v_1$	$((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$

Notice that $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ is a subformula of $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$. In general, *every formula counts as a subformula of itself*. A subformula of a formula φ that is not φ itself is call a *proper* subformula of φ.

Definition (11) states that subformulas of a formula φ are consecutive sequences of symbols appearing in φ. But that doesn't mean that *every* consecutive sequence of symbols appearing in φ is a subformula of φ; the sequence has to be a formula itself. For example, $\neg v_{201}\wedge$ is a sequence of symbols occurring in the formula $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$. But $\neg v_{201}\wedge$ is not a subformula of $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ (since it is not a formula).

(13) SURPRISE QUIZ: Why isn't $v_{201} \wedge v_{39}$ listed in (12) as one of the subformulas of $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$? Isn't it true that $v_{201} \wedge v_{39}$ is a consecutive sequence of symbols in $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ that is itself a formula?

The quiz is tricky; don't feel badly if it stumped you. The answer is that $v_{201} \wedge v_{39}$ is not a subformula of $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ because it is not a formula. It is not a formula because it is missing the parentheses that must surround conjunctions. Parentheses matter, like we said.

3.5 Construction tables

The steps described in (9) to make $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ can be represented in a kind of table that we'll call a *construction table* for $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$. Here is such a table. It should be read from the bottom up.

(14) Construction table for $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$:

<table>
<tr><th colspan="3">subformulas</th><th>clause</th></tr>
<tr><td colspan="3">$((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$</td><td>$(8)d$</td></tr>
<tr><td colspan="2">$(\neg\neg v_{201} \wedge v_{39})$</td><td></td><td>$(8)c$</td></tr>
<tr><td>$\neg\neg v_{201}$</td><td></td><td></td><td>$(8)b$</td></tr>
<tr><td>$\neg v_{201}$</td><td></td><td>$\neg v_1$</td><td>$(8)b$</td></tr>
<tr><td>v_{201}</td><td>v_{39}</td><td>v_1</td><td>$(8)a$</td></tr>
</table>

The clauses from Definition (8) are displayed at the right of the table. The other entries are the subformulas of $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ that are constructed along the way. We note that *all the subformulas of the top formula in a construction table appear in the body of the table.* Also, except for vertical lines and clause labels, *nothing other than subformulas of the top formula appear in a given construction table*.

Here are some more tables. The first corresponds to the steps described in (10).

(15) Construction table for $\neg((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$:

<table>
<tr><th colspan="4">subformulas</th><th>clause</th></tr>
<tr><td colspan="4">$\neg((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$</td><td>$(8)b$</td></tr>
<tr><td colspan="4">$((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$</td><td>$(8)e$</td></tr>
<tr><td colspan="2"></td><td colspan="2">$(v_6 \vee v_7)$</td><td>$(8)d$</td></tr>
<tr><td colspan="2">$(v_3 \wedge v_5)$</td><td></td><td></td><td>$(8)c$</td></tr>
<tr><td>v_3</td><td>v_5</td><td>v_6</td><td>v_7</td><td>$(8)a$</td></tr>
</table>

(16) Construction table for $((v_1 \rightarrow v_2) \rightarrow \neg v_1)$:

subformulas			*clause*
$((v_1 \to v_2) \to \neg v_1)$			$(8)c$
$(v_1 \to v_2)$			$(8)e$
		$\neg v_1$	$(8)b$
v_1	v_2	v_1	$(8)a$

Notice that we could reorganize Table (16) somewhat, and write it as follows.

(17)

subformulas			*clause*
$((v_1 \to v_2) \to \neg v_1)$			$(8)c$
		$\neg v_1$	$(8)b$
$(v_1 \to v_2)$			$(8)e$
v_1	v_2	v_1	$(8)a$

That both (16) and (17) describe the construction of $((v_1 \to v_2) \to \neg v_1)$ shows there is not a unique construction table for a formula; there may be more than one. It doesn't much matter which construction table we build for a formula, however, because of the following fact. *Every construction table for a given formula exhibits the same set of subformulas for that formula.* To illustrate, the two tables (16) and (17) for $((v_1 \to v_2) \to \neg v_1)$ both list $v_1, v_2, (v_1 \to v_2), \neg v_2$, and $((v_1 \to v_2) \to \neg v_1)$ as its subformulas. The fact stated above in italics is an immediate consequence of the other italicized facts stated earlier. We can prove them all rigorously only by giving a formal definition of *construction table*. It's not worth the bother; you can just trust us in this matter.

3.6 Types of formulas

A formula that consists of a single variable (without any connectives) is called *atomic*. For example, the formula v_{41} is atomic. Atomic formulas correspond to declarative sentences (e.g., of English) whose internal structure is not dissected by our logic. All the other formulas are called *nonatomic*. Nonatomic formulas include at least one connective.

The *principal connective* of a nonatomic formula φ is the connective that is inserted last in the construction table for φ. This connective is unique; it doesn't depend on which particular construction table you build for a formula. To illustrate, Table (14) shows that the principal connective of $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ is the $\vee$. Table (16) shows that the principal connective of $((v_1 \rightarrow v_2) \rightarrow \neg v_1)$ is the rightmost $\rightarrow$. Table (15) shows that the principal connective of $\neg((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$ is the $\neg$. The principal connective of $\neg\neg\neg v_{98}$ is the leftmost $\neg$. There is no principal connective in the atomic formula v_{101} (because there are no connectives at all). We classify nonatomic formulas according to their principal connective. In particular:

A nonatomic formula whose principal connective is $\neg$ is called a *negation*. Such a formula can be represented as having the form $\neg\varphi$.

A nonatomic formula whose principal connective is $\wedge$ is called a *conjunction*. Such a formula can be represented as having the form $(\varphi \wedge \psi)$. The subformulas φ, ψ are called the *conjuncts* of this conjunction.

A nonatomic formula whose principal connective is $\vee$ is called a *disjunction*. Such a formula can be represented as having the form $(\varphi \vee \psi)$. The subformulas φ, ψ are called the *disjuncts* of this disjunction.

A nonatomic formula whose principal connective is $\rightarrow$ is called a *conditional*. Such a formula can be represented as having the form $(\varphi \rightarrow \psi)$. The subformula φ is called the *left hand side* of the conditional, and the subformula ψ is called the *right hand side*.

A nonatomic formula whose principal connective is $\leftrightarrow$ is called a *biconditional*. Such a formula can be represented as having the form $(\varphi \leftrightarrow \psi)$. We again use *left hand side* and *right hand side* to denote the subformulas φ, ψ.

Thus, $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ is a disjunction with disjuncts $(\neg\neg v_{201} \wedge v_{39})$ and $\neg v_1$. The leftmost disjunct is a conjunction with conjuncts $\neg\neg v_{201}$ and v_{39}. The leftmost conjunct of the latter conjunction is a negation (of a negation). The formula, $((v_1 \rightarrow v_2) \rightarrow \neg v_1)$ is a conditional with left hand side $(v_1 \rightarrow v_2)$ and right hand side $\neg v_1$. The left hand side is itself a conditional whereas the right hand side is a negation. The formula $\neg((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$ is the negation of $((v_3 \wedge v_5) \leftrightarrow (v_6 \vee v_7))$, which is a biconditional with left hand side the conjunction $(v_3 \wedge v_5)$ and right hand side the disjunction $(v_6 \vee v_7)$.

We say that the conjuncts of a conjunction are its "principal subformulas," and likewise for the disjuncts of a disjunction, etc. Let's record this useful terminology.

(18) DEFINITION: Let a formula $\varphi \in \mathcal{L}$ be given. We define as follows the *principal subformulas* of φ.

(a) If φ is atomic then φ has no principal subformulas.

(b) If φ is the negation $\neg\psi$ then ψ is the principal subformula of φ (there are no others).

(c) If φ is the conjunction $(\psi \wedge \chi)$, disjunction $(\psi \vee \chi)$, conditional $(\psi \rightarrow \chi)$, or biconditional $(\psi \leftrightarrow \chi)$ then ψ and χ are the principal subformulas of φ (there are no others).

Here is a nice fact that we'll use in Chapter 7. Suppose that $\varphi \in \mathcal{L}$ is not atomic.[6] Then φ has a principal connective. By thinking about some examples, you should be able to see that the only consecutive sequence of symbols in φ that is itself a formula and includes φ's principal connective is φ itself. To illustrate, let φ be $((p \vee q) \rightarrow (\neg q \rightarrow r))$. The principal connective is the leftmost $\rightarrow$. There is no proper subformula of φ that includes this leftmost $\rightarrow$.[7] We formulate our insight:

(19) FACT: Suppose that $\varphi \in \mathcal{L}$ is not atomic. Then no proper subformula of φ includes the principal connective of φ.

[6] This sentence means: let a formula φ be given, and suppose that φ is not atomic.

[7] Reminder: A "proper" subformula of a formula φ is any subformula of φ other than φ itself.

3.7 Diction

Each of the symbols of $\mathcal{L}$ has a name in English. Thus, v_7 is known (affectionately) as "vee seven," and likewise for the other sentential variables. We'll use the word "tilde" to name the symbol $\neg$ (even though "tilde" applies better to another common notation for negation, namely $\sim$). The word "wedge" names the symbol $\wedge$, and "vee" names $\vee$. We use "arrow" and "double arrow" for $\rightarrow$ and $\leftrightarrow$, respectively. The name for (is "left parenthesis," and the name for) is "right parenthesis" (perhaps you knew this.) So, to pronounce a formula in English, you can revert to naming its symbols. For example, $(v_7 \wedge \neg v_8)$ is pronounced "left parenthesis, vee seven, wedge, tilde, vee eight right parenthesis."

This makes for some riveting dinner conversation, but even spicier remarks result from pronouncing the connectives using English phrases. For this purpose, use a negative construction (like "not") for $\neg$, a conjunctive construction (like "and") for $\wedge$, a disjunctive construction (like "or") for $\vee$, a conditional construction (like "if–then–") for $\rightarrow$, and a biconditional construction (like "if and only if") for $\leftrightarrow$. A few examples will communicate how it goes.

You can pronounce $(\neg v_1 \wedge v_2)$ as "Both not vee one and vee two." Notice that without the word "both" your sentence would be ambiguous between $(\neg v_1 \wedge v_2)$ and $\neg(v_1 \wedge v_2)$. The word "both" serves to mark the placement of the parenthesis. To pronounce $\neg(v_1 \wedge v_2)$ you say "Not both vee one and vee two." For $(v_1 \wedge \neg v_2)$ you can simply say "Vee one and not vee two" (there is no ambiguity). Likewise, $(\neg v_1 \vee v_2)$ should be pronounced "Either not vee one or vee two." To pronounce $\neg(v_1 \vee v_2)$ you say "Not either vee one or vee two." For $(v_1 \vee \neg v_2)$ you can say "Vee one or not vee two."

You can pronounce $(v_2 \rightarrow \neg(v_2 \wedge v_3))$ as "if vee two then not both vee two and vee three." You can pronounce $(v_2 \leftrightarrow (\neg v_2 \rightarrow v_3)$ as "vee two if and only if if not vee two then vee three."

No doubt you get it, so we won't go on with further examples. This way of pronouncing formulas of $\mathcal{L}$ is entirely unofficial anyway. Remember, the whole point of constructing the artificial language $\mathcal{L}$ is to avoid the tangled syntax of natural languages like English, whose sentences are hard to interpret mathematically. The only virtue of such loose talk at this point is to suggest the kind of meaning we have in mind for the wedge, vee, and so forth. Once they are assigned their official meanings, we'll return to the question of what English expressions express them.

Can you do us a favor now, though? Please don't use the word "implies" to pronounce the arrow. That is, don't pronounce $(v_1 \rightarrow v_2)$ as "vee one implies vee two." We'll see later that whatever the arrow means in logic, it doesn't mean "implies."

3.8 Abbreviation

Aren't you getting tired of the subscripts on the sentential variables? Often it is simpler to pretend that the list of variables include the letters from p to the end of the alphabet. Henceforth, we'll often use these letters in place of v's when presenting formulas. Thus, instead of writing $(v_2 \rightarrow \neg v_{91})$ we'll just write $(p \rightarrow \neg q)$. In this sense, letters like p and q are abbreviations for particular variables. The official vocabulary for $\mathcal{L}$ doesn't have these letters. They just serve to refer to genuine variables. (We could have included them at the beginning of our list of sentential variables, but that would have been unbearably ugly.)

Another abbreviation consists in eliminating outer pairs of parentheses. Thus, we often write $p \wedge q$ in place of $(p \wedge q)$. We can do this because it is clear where to put the parentheses in $p \wedge q$ to retrieve the legal formula $(p \wedge q)$.[8] We also allow ourselves to dispense with parentheses when conjunctions of a conjunction are conjunctions. Thus, we may write $(p \rightarrow q) \wedge q \wedge (r \leftrightarrow q)$ in place of $((p \rightarrow q) \wedge q) \wedge (r \leftrightarrow q)$ or $(p \rightarrow q) \wedge (q \wedge (r \leftrightarrow q))$. In this case there is ambiguity about which of the latter two formulas is intended by $(p \rightarrow q) \wedge q \wedge (r \leftrightarrow q)$. But we'll see in the next chapter that the two possibilities come to the same thing so it is often more convenient to allow the ambiguity. Likewise, we often write $(p \rightarrow q) \vee q \vee (r \leftrightarrow q)$ in place of either $((p \rightarrow q) \vee q) \vee (r \leftrightarrow q)$ or $(p \rightarrow q) \vee (q \vee (r \leftrightarrow q))$.

[8]Be sure to put the parentheses back before searching for subformulas of a given formula. Abbreviated or not, $((\neg\neg v_{201} \wedge v_{39}) \vee \neg v_1)$ does not count $v_{201} \wedge v_{39}$ among its subformulas. See the Surprise Quiz (13).

In other logic books, you are invited to memorize a precedence order on connectives. Higher precedence for a connective is interpreted as placing it higher in the construction table for a formula. For example, the arrow is typically given higher precedence than the wedge. This allows the formula $(p \wedge q) \rightarrow r$ to be written unambiguously as $p \wedge q \rightarrow r$. The latter expression cannot then be interpreted as $p \wedge (q \rightarrow r)$. Some folks accept such precedence schemes in exchange for eliminating parentheses. In contrast, we will live with a few extra parentheses in order to lighten your memory load. There is no precedence among connectives; we are logical egalitarians.

Finally, while we're criticizing the practices of other authors, here are some notations for connectives that you may find in place of ours.

ours	theirs			
$\neg p$	$\sim p$	$\bar{p}$		
$p \wedge q$	$p \,\&\, q$	pq	$p.q$	$p \cdot q$
$p \vee q$	$p \mid q$	$p + q$		
$p \rightarrow q$	$p \supset q$	$p \Rightarrow q$		
$p \leftrightarrow q$	$p \equiv q$	$p \Leftrightarrow q$		

The alternative notation is aesthetically challenged, to say the least. Our choices are currently the most common.

3.9 More Greek

When defining the formulas of $\mathcal{L}$ in Section 3.3, we had recourse to Greek letters like φ, ψ, and χ.[9] They were used to represent arbitrary formulas that served as building blocks in Definition (8). The same machinery will be needed for many other purposes, so we take the present opportunity to clarify its use.

If we write the string of characters "$(\varphi \vee \psi)$" we are referring to a disjunction with disjuncts denoted by φ and ψ. The latter may represent any formulas whatsoever, even the same one (unless other conditions are stated). Thus, $(\varphi \vee \psi)$ stands ambiguously for any of the formulas $((p \vee q) \vee r)$, $((\neg s \wedge \neg r) \vee (r \rightarrow \neg s))$,

[9]To remind you, the letters are pronounced "figh," "sigh," and "kigh," respectively.

$(q \vee q)$, etc. Similarly, $((\varphi \to \psi) \wedge \chi)$ stands ambiguously for any of $(((r \wedge q) \to \neg p) \wedge \neg s)$, $((p \to p) \wedge p)$, etc. On the other hand, if we repeat a Greek letter in such an expression, then the formulas denoted are meant to be the same. Thus, $(\varphi \wedge \varphi)$ can denote $((q \to r) \wedge (q \to r))$ but not $((q \to r) \wedge (r \to q))$. Each greek letter has just one interpretation in a given expression.

Such notation will be used in a free-swinging way. In particular, we'll allow ourselves to talk of "the formula $(\varphi \leftrightarrow \psi)$" instead of the more exact "a formula of form $(\varphi \leftrightarrow \psi)$." We'll also occasionally drop parentheses according to the abbreviations discussed in Section 3.8. Thus, $\varphi \wedge \psi$ will stand for any conjunction.

(20) EXERCISE: Exhibit construction tables for the following formulas, and describe them as we did in Section 3.6.

(a) $\neg(p \wedge q) \leftrightarrow r$

(b) $p \to (q \vee (q \to \neg r))$

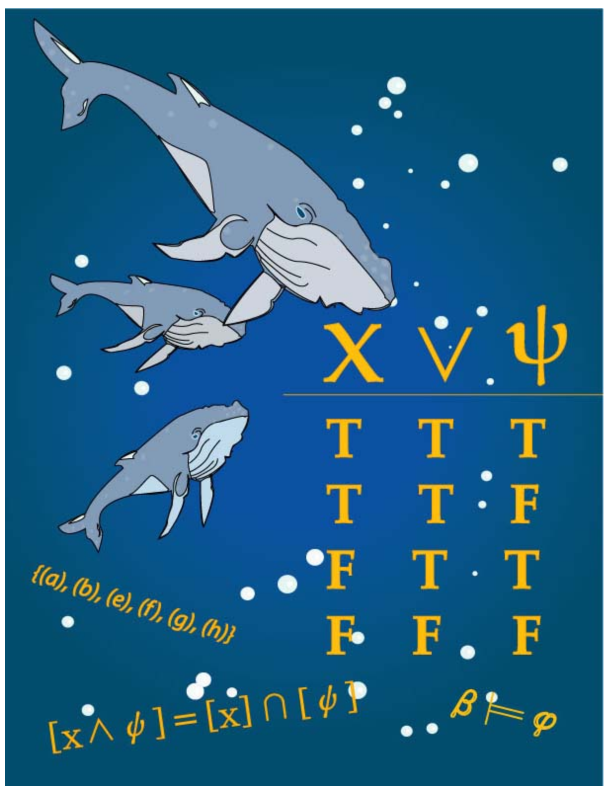
(c) $\neg(\neg p \wedge (p \vee (q \to \neg p)))$

(d) $\neg\neg p \wedge (p \vee (q \to \neg p))$

(21) EXERCISE: Write some formulas that are referred to by $(\varphi \to \psi) \wedge (\psi \to \varphi)$. Is $(p \to q) \wedge (p \to p)$ one of them?

Chapter 4

Meaning



4.1 The semantics project

"But what does it all mean?"

We take your question to be about the formulas of $\mathcal{L}$, the language of sentential logic. You want to know how the formulas are to be interpreted, what determines whether a given formula is true or false, which of them have the same meaning, and so forth. The present chapter responds to these questions, and thus explains the *semantics* of $\mathcal{L}$.[1]

Before getting embroiled in $\mathcal{L}$, let us reflect on what a semantic theory of English would look like. One appealing conception is based on two sets, S and M. S holds all the sequences of English words that can be written or spoken. M holds all the meanings people might want to express. What a sentence in S means depends on the situation in which it is written or spoken, so our understanding of English is represented by our ability to map members of S into the contextually right members of M. Some sequences will map onto a single meaning. For example, the semantics of English dictate that an instance of the sentence "The baby is finally asleep" is mapped into the meaning that often enters the mind of the new parents around 2:00 a.m. The situation determines which baby and which parents and which time is relevant. Instances of the string "be bop be do" is mapped to nothing in M. Even genuine English sentences need not be mapped to unique members of M. For example, "The D.A. loves smoking guns" is often mapped into two meanings, and "Beauty is eternity's self-embrace" seems not be mapped into any meaning at all.[2]

Such an attractive picture of semantics is worth pursuing, and it *has* been pursued in such works as [16, 58, 63]. One of the challenges in formulating the theory is giving substance to the class M. What exactly is a meaning? Facing this challenge requires formalizing the kind of thing that gets "meant" by sentences of natural language. Another challenge is giving a precise character-

[1]According to the dictionary [1], semantics is "the study or science of meaning in language."

[2]Conversely, members of M that are hit by no sentence are "ineffable" in English. The meanings expressed in ballads sung by whales might be ineffable. (Our favorite whale ballad is *A tuna for my baby*.) Perhaps there are meanings expressed in certain natural languages (e.g., Mohawk) that are ineffable in English. The matter has been richly debated; see Kay & Kempton (1984) [59].

ization of what a situation is and what features of a situation are relevant in a particular instances. A different path is to turn one's back on the problem of meaning in natural language, and try instead to provide a useful semantics for some artificial language. This unglorious route is taken in the present chapter, and the artificial language in question is $\mathcal{L}$.

Semantics is relatively painless for $\mathcal{L}$ because we are able to specify in advance the class of meanings. It then remains only to say which meaning is assigned to which member of $\mathcal{L}$. Exactly one meaning is associated with a given formula since $\mathcal{L}$ is free from ambiguity and meaninglessness. Its interpretation is also *independent of context* inasmuch as a given formula means the same thing no matter what other formulas are written nearby. In Chapter 3 we introduced the members of $\mathcal{L}$. In Section 4.3 below we introduce the class of meanings. Then we'll correlate the two. But there is an important preliminary step. Meanings will be defined as sets of mappings of the sentential variables into truth values. These mappings need first to be given a name and explained

4.2 Truth-assignments

Our goal is to attach meanings to formulas. This requires explaining meanings (we already explained formulas in the previous chapter). To explain meanings, we require the fundamental concept of "truth-assignment." This concept is introduced in the present section.

4.2.1 Variables mapped to truth and falsity

To get started, let us hypothesize the existence of two abstract objects called *the true* and *the false*. They are the "truth values." We'll symbolize them by T and F. To make sure there is no misunderstanding at this early stage, let us record:

(1) DEFINITION: The set of *truth values* is $\{\mathsf{T}, \mathsf{F}\}$, consisting of the true and the false.

Now we ask you to recall the idea of a "mapping" from one set to another.[3] This concept allows us to present a key idea of logic.

(2) DEFINITION: Any mapping from the set of sentential variables of $\mathcal{L}$ to the set $\{\mathsf{T}, \mathsf{F}\}$ of truth values is a *truth-assignment*.

To make sure that you are clear about the definition, we will spell it out in a simple case. Suppose that the number of sentential variables in $\mathcal{L}$ is 3. (That is, suppose you fixed n to be 3 in Section 3.2.) We'll use p, q, r to denote these variables. Then one truth-assignment assigns T to all three variables. Another assigns F to all of them. Yet another assigns F to p and T to q and r. In all, there are 8 truth-assignments when $n = 3$. Let's list them in a table.

(3)

	p	q	r
(a)	T	T	T
(b)	T	T	F
(c)	T	F	T
(d)	T	F	F
(e)	F	T	T
(f)	F	T	F
(g)	F	F	T
(h)	F	F	F

(All 8 truth-assignments for 3 variables, p, q, r.)

For example, the truth-assignment labeled (g) in (3) assigns F to p and q, and T to r.

Why did it turn out that there are 8 truth-assignments for 3 variables? Well, p can be either T or F, which makes two possibilities. The same holds for q, and its two possibilities are independent of those for p. Same thing for r; it can be set to T or F independently of the choices for p and q. This makes $2 \times 2 \times 2 = 8$ ways of assigning T and F to the three variables p, q, r. Each way is a truth-assignment, so there are eight of them. Extending this reasoning, we see:

(4) FACT: If n is the number of sentential variables in $\mathcal{L}$ then there are 2^n truth-assignments.

[3] See Section 2.10.

We may now name the set containing all of the truth-assignments.

(5) DEFINITION: The set of all truth-assignments for our language $\mathcal{L}$ is denoted by the symbol TrAs.

If $n = 3$ then TrAs is the set of eight truth-assignments appearing in Table (3).

4.2.2 Truth-assignments extended to $\mathcal{L}$

The French say that "better" is the enemy of "good." Defying this aphorism, we shall now make the concept of truth-assignment even better. In its original form (defined above), a truth-assignment maps every variable to a truth value. When we're done with the present subsection, they will do this and more. They will map every formula (including variables) to truth values. Once this is accomplished, truth-assignments will be ready for their role in the definition of meaning.

Another recursive definition. Do you remember what an atomic formula is? In Section 3.6 they were defined to be the sentential variables, in other words, the formulas with no connectives. A given truth-assignment maps each atomic formula into one of T and F. It does not apply to nonatomic formulas like $p \vee (q \rightarrow \neg p)$. The next definition fixes this. Given a truth-assignment α, we define its "extension" to all of $\mathcal{L}$, and denote the extension by $\overline{\alpha}$. The definition is recursive.[4] It starts by specifying the truth value assigned to a variable like v_i. [This truth value is denoted $\overline{\alpha}(v_i)$.] Then the definition *supposes* that $\overline{\alpha}$ has assigned truth values to some formulas, and goes on to say what truth value $\overline{\alpha}$ assigns to more complicated formulas. Let's do it.

(6) DEFINITION: Suppose that a truth-assignment α and a formula φ are given. φ is either atomic, a negation, a conjunction, a disjunction, a conditional, or a biconditional. We define $\overline{\alpha}(\varphi)$ (the value of the mapping $\overline{\alpha}$ on the input φ) in all these cases.

[4] You encountered a recursive definition earlier, in Section 3.3.

(a) Suppose that φ is the atomic formula v_i. Then $\overline{\alpha}(\varphi) = \alpha(v_i)$.

(b) Suppose that φ is the negation $\neg\psi$, and that $\overline{\alpha}(\psi)$ has already been defined. Then $\overline{\alpha}(\varphi) = \mathsf{T}$ if $\overline{\alpha}(\psi) = \mathsf{F}$, and $\overline{\alpha}(\varphi) = \mathsf{F}$ if $\overline{\alpha}(\psi) = \mathsf{T}$.

(c) Suppose that φ is the conjunction $\chi \wedge \psi$, and that $\overline{\alpha}(\chi)$ and $\overline{\alpha}(\psi)$ have already been defined. Then $\overline{\alpha}(\varphi) = \mathsf{T}$ just in case $\overline{\alpha}(\chi) = \mathsf{T}$ and $\overline{\alpha}(\psi) = \mathsf{T}$. Otherwise, $\overline{\alpha}(\varphi) = \mathsf{F}$.

(d) Suppose that φ is the disjunction $\chi \vee \psi$, and that $\overline{\alpha}(\chi)$ and $\overline{\alpha}(\psi)$ have already been defined. Then $\overline{\alpha}(\varphi) = \mathsf{F}$ just in case $\overline{\alpha}(\chi) = \mathsf{F}$ and $\overline{\alpha}(\psi) = \mathsf{F}$. Otherwise, $\overline{\alpha}(\varphi) = \mathsf{T}$.

(e) Suppose that φ is the conditional $\chi \rightarrow \psi$, and that $\overline{\alpha}(\chi)$ and $\overline{\alpha}(\psi)$ have already been defined. Then $\overline{\alpha}(\varphi) = \mathsf{F}$ just in case $\overline{\alpha}(\chi) = \mathsf{T}$ and $\overline{\alpha}(\psi) = \mathsf{F}$. Otherwise, $\overline{\alpha}(\varphi) = \mathsf{T}$.

(f) Suppose that φ is the biconditional $\chi \leftrightarrow \psi$, and that $\overline{\alpha}(\chi)$ and $\overline{\alpha}(\psi)$ have already been defined. Then $\overline{\alpha}(\varphi) = \mathsf{T}$ just in case $\overline{\alpha}(\chi) = \overline{\alpha}(\psi)$. Otherwise, $\overline{\alpha}(\varphi) = \mathsf{F}$.

We comment on each clause of Definition (6). Our plan was that $\overline{\alpha}$ be an extension of α. That is, $\overline{\alpha}$ should never contradict α, but rather agree with what α says while saying more. It is clause (6)a that guarantees success in this plan. $\overline{\alpha}$ is an extension of α because $\overline{\alpha}$ gives the same result as α when applied to sentential variables (which are the only thing to which α applies).

Clause (6)b says that $\neg$ has the effect of switching truth values: $\overline{\alpha}(\neg\psi)$ is T or F as $\overline{\alpha}(\psi)$ is F or T respectively. The tilde thus expresses a basic form of negation.

Clause (6)c imposes a conjunctive reading on $\wedge$. Given a formula of form $\chi \wedge \psi$ and a truth-assignment α, $\overline{\alpha}$ assigns T to the formula just in case it assigns T to χ *and* it assigns T to ψ.

Similarly, clause (6)d imposes a disjunctive reading on $\vee$. Given a formula of form $\chi \vee \psi$ and a truth-assignment α, $\overline{\alpha}$ assigns T to the formula just in case *either* it assigns T to χ *or* it assigns T to ψ. You'll need to keep in mind that the "either ... or ..." just invoked is *inclusive* in meaning. That is, if $\overline{\alpha}$ assigns T both to χ and to ψ, then it assigns T to their disjunction $\chi \vee \psi$. An *exclusive* reading would assign truth to a disjunction just in case exactly one of the disjuncts is true. Perhaps "or" is interpreted exclusively in the sentence: "For dessert, you may have either cake or ice cream." Exclusive readings in English are often signaled by the tag "but not both!". To repeat: in Sentential Logic, the disjunction $\chi \vee \psi$ is understood inclusively.

Formulas of the form $\chi \rightarrow \psi$ are supposed to express the idea that if χ is true than so is ψ. Clause (6)e cashes in this idea by setting $\overline{\alpha}(\chi \rightarrow \psi)$ to F only in case $\overline{\alpha}(\chi)$ is T and $\overline{\alpha}(\psi)$ is F. In the three other cases, $\overline{\alpha}$ sets $\chi \rightarrow \psi$ to T. Is this a proper rendering of *if–then–*? The question will be cause for much anguish in Chapter 8.

Clause (6)f makes biconditionals express the claim that the left hand side and right hand side have the same truth value. Thus $\overline{\alpha}(\chi \leftrightarrow \psi) = \mathsf{T}$ just in case either $\overline{\alpha}(\chi) = \mathsf{T}$ and $\overline{\alpha}(\psi) = \mathsf{T}$, or $\overline{\alpha}(\chi) = \mathsf{F}$ and $\overline{\alpha}(\psi) = \mathsf{F}$.

"Do we really have to remember all this?"

Yes, you do, but it's not as hard as it looks. Studying the examples and working the exercises should just about suffice for assimilating Definition (6). In case of doubt, try to remember the rough meaning of each connective, namely $\neg$ for "not," $\wedge$ for "and," $\vee$ for "or," $\rightarrow$ for "if ... then ...," and $\leftrightarrow$ for "if and only if." Then most of Definition (6) makes intuitive sense. For example, a truth-assignment makes $\chi \wedge \psi$ true just in case it makes both χ and ψ true. The other connectives can be thought of similarly except for two qualifications. First, $\chi \vee \psi$ (that is, "χ or ψ") is true in a truth-assignment even if *both* χ and ψ are true. So, $\chi \vee \psi$ should not be understood as "Either χ or ψ but not both" as in "Either Smith will win the race or Jones will win the race but not both." (As mentioned above, the latter interpretation of "or" is exclusive in contrast to the inclusive interpretation that we have reserved for $\vee$.) Second, $\chi \rightarrow \psi$ (that is " If χ then ψ") is *false* in just one circumstance, namely, if the truth-assignment in question makes χ true and ψ false. The other three ways of assigning truth and falsity to χ and ψ make $\chi \rightarrow \psi$ true. It makes intuitive sense to declare "If χ then ψ") false if χ is true and ψ is false. (Example: "If the computer is on then Bob is working" is false if the computer is on but Bob's isn't working.) All you need to remember is the less intuitive stipulation that every other combination of truth values makes the conditional true. In our experience, students tend to forget the semantics of $\rightarrow$; the other connectives are remembered better.

Applying the definition to arbitrary formulas. The six clauses of Definition (6) work recursively to assign a unique truth value to any $\varphi \in \mathcal{L}$, given a truth-assignment α. For example, consider truth-assignment (c) in Table (3), namely, the assignment of T to p and r, and F to q. What is $\overline{(c)}(\neg r \vee (p \wedge q))$?[5] Well, by (6)a, $\overline{(c)}(r) = (c)(r) = \mathsf{T}$, $\overline{(c)}(p) = (c)(p) = \mathsf{T}$, and $\overline{(c)}(q) = (c)(q) = \mathsf{F}$, Hence, by (6)b, $\overline{(c)}(\neg r) = \mathsf{F}$, and by (6)c, $\overline{(c)}(p \wedge q) = \mathsf{F}$. So by (6)d, $\overline{(c)}(\neg r \vee (p \wedge q)) = \mathsf{F}$. This reasoning can be summarized by adding truth values to the subformulas appearing in the construction table for $\neg r \vee (p \wedge q)$. Read the following table from the bottom up. Truth values are added in brackets after each subformula.

(7) Construction table for $\neg r \vee (p \wedge q)$, augmented by truth values due to $\overline{(c)}$:

<table>
<tr><td colspan="3">$\neg r \vee (p \wedge q)$ [F]</td></tr>
<tr><td></td><td colspan="2">$(p \wedge q)$ [F]</td></tr>
<tr><td>$\neg r$ [F]</td><td></td><td></td></tr>
<tr><td>r [T]</td><td>p [T]</td><td>q [F]</td></tr>
</table>

From this example you see that Definition (6) does its work by climbing up construction tables. First it assigns truth values to the formulas (variables) at the bottom row, then to the next-to-last row, and so forth up to the top.

Here is another example. Consider truth-assignment (d) in Table (3), namely, the assignment of F to q and r, and T to p. What is $\overline{(d)}((p \vee q) \rightarrow p)$? By (6)a, $\overline{(d)}(p) = (d)(p) = \mathsf{T}$, $\overline{(d)}(q) = (d)(q) = \mathsf{F}$. So by (6)d, $\overline{(d)}(p \vee q) = \mathsf{T}$, and by (6)e, $\overline{(d)}((p \vee q) \rightarrow p) = \mathsf{T}$. The corresponding augmented construction table is:

(8) Construction table for $(p \vee q) \rightarrow p$, augmented by truth values due to $\overline{(d)}$:

[5]If you're confused by this question, it may help to parse the expression $\overline{(c)}(\neg r \vee (p \wedge q))$. You know that c is a truth-assignment, hence, a function from variables to truth values. Its extension $\overline{(c)}$ is a function from all of $\mathcal{L}$ to truth values. So $\overline{(c)}(\neg r \vee (p \wedge q))$ is the application of this latter function to a specific formula. The result is a truth value. Therefore $\overline{(c)}(\neg r \vee (p \wedge q))$ denotes a truth value, in other words, either T or F. The question we posed is: Which of T, F is denoted by $\overline{(c)}(\neg r \vee (p \wedge q))$?

<table>
<tr><td colspan="3">$(p \vee q) \rightarrow p$ [T]</td></tr>
<tr><td colspan="2">$(p \vee q)$ [T]</td><td></td></tr>
<tr><td>p [T]</td><td>q [F]</td><td>p [T]</td></tr>
</table>

The foregoing examples illustrate the claim made above: Definition (6) extends any truth-assignment α into a unique mapping $\overline{\alpha}$ of each formula of $\mathcal{L}$ into exactly one truth value.

Look again at the augmented construction table (7). If you tug on the bottom it will zip up like a window shade into the following.

(9)

$\neg$	r	$\vee$	$(p$	$\wedge$	$q)$
F	T	F	T	F	F

This condensed display summarizes the computation of $\overline{(c)}(\neg r \vee (p \wedge q)$ if you build it from small subformulas to big ones ("inside out"). You start with the variables (no parentheses), and mark them with their truth values according to $\overline{(c)}$. Then proceed to the larger subformulas that can be built from the variables, namely $\neg p$ and $(p \wedge q)$. Once their principal connectives are marked with truth values, you proceed to the next largest subformula, which happens to be the whole formula $\neg r \vee (p \wedge q)$. Its truth value [according to $\overline{(c)}$] is written below its principal connective, namely, the $\vee$. Similarly, Table (8) zips up into:

(10)

$(p$	$\vee$	$q)$	$\rightarrow$	p
T	T	F	T	T

If you've understood these ideas, you should be able to calculate the truth value of any formula according to the extension of any truth-assignment. A little practice won't hurt.

(11) EXERCISE: Calculate the truth values of the following formulas according to the (extensions of) truth-assignments (a), (b), (g), and (h) in Table (3).

$r \wedge (r \rightarrow \neg(p \wedge q))$

$(p \rightarrow (p \rightarrow p)) \rightarrow p$

$\neg(q \vee \neg(p \vee q))$

$(p \vee q) \leftrightarrow (q \wedge \neg r)$

4.2.3 Further remarks and more notation.

Recall that when $\mathcal{L}$ has n variables, there are 2^n truth-assignments. How many extended truth-assignments are there for n variables? That is, how many members are there in the set $\{\overline{\alpha} \mid \alpha \in \mathsf{TrAs}\}$? This is a trick question. Think for a minute.

Of course, the answer is 2^n again. Otherwise, extensions of truth-assignments wouldn't be unique, and they are.

Notice that $\overline{\alpha}$ only cares about truth values. For example, if $\alpha(p) = \alpha(r)$ and $\alpha(q) = \alpha(s)$ then $\overline{\alpha}(p \wedge q) = \overline{\alpha}(r \wedge s)$. Specifically, if $\alpha(p) = \alpha(r) = \mathsf{T}$ and $\alpha(q) = \alpha(s) = \mathsf{F}$, then $\overline{\alpha}(p \wedge q) = \overline{\alpha}(r \wedge s) = \mathsf{F}$, and similarly for the other possible combinations of truth values. We can put the matter this way: only the truth values of p, q contribute to computing $\overline{\alpha}(p \wedge q)$. In this sense, the logical connectives are *truth functional*.

The truth functionality of connectives can be expressed slightly differently. Suppose that truth-assignments α and β agree about the variables p and q, that is, $\alpha(p) = \beta(p)$ and $\alpha(q) = \beta(q)$. Then you can see from Definition (6)c that $\overline{\alpha}(p \wedge q) = \overline{\beta}(p \wedge q)$. Again, only the truth values of p, q contribute to computing $\overline{\alpha}(p \wedge q)$. By examining the other clauses of Definition (6) you'll recognize the following, more general points.

(12) FACT: Let α and β be two truth-assignments.

(a) Suppose that α and β agree about the variables $u_1 \cdots u_k$, that is, $\alpha(u_1) = \beta(u_1) \cdots \alpha(u_k) = \beta(u_k)$. Suppose that variables appearing in $\varphi \in \mathcal{L}$ are a subset of $\{u_1 \cdots u_k\}$ (that is, no variable outside the list $u_1 \cdots u_k$ occurs in φ). Then $\overline{\alpha}$ and $\overline{\beta}$ agree about φ, that is, $\overline{\alpha}(\varphi) = \overline{\beta}(\varphi)$.

(b) Suppose that $\overline{\alpha}$ and $\overline{\beta}$ agree about the formulas φ, ψ, that is, $\overline{\alpha}(\varphi) = \overline{\beta}(\varphi)$ and $\overline{\alpha}(\psi) = \overline{\beta}(\psi)$. Then $\overline{\alpha}$ and $\overline{\beta}$ agree about $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$,

$\varphi \to \psi$ and $\varphi \leftrightarrow \psi$. That is:

i. $\overline{\alpha}(\neg\varphi) = \overline{\beta}(\neg\varphi)$.
ii. $\overline{\alpha}(\varphi \wedge \psi) = \overline{\beta}(\varphi \wedge \psi)$.
iii. $\overline{\alpha}(\varphi \vee \psi) = \overline{\beta}(\varphi \vee \psi)$.
iv. $\overline{\alpha}(\varphi \to \psi) = \overline{\beta}(\varphi \to \psi)$.
v. $\overline{\alpha}(\varphi \leftrightarrow \psi) = \overline{\beta}(\varphi \leftrightarrow \psi)$.

The semantics of $\mathcal{L}$ have another neat property that's worth mentioning. Consider how we use Definition (6) to augment the construction table for a formula φ. When deciding whether to assign T versus F to a nonatomic subformula ψ of φ, all that matters are the truth-values assigned to the principal subformulas of ψ (or to the sole, principal subformula in case ψ is a negation).[6] No other subformula of φ gets involved. The point is illustrated by the following augmented construction table for $\neg(\neg r \vee (p \wedge q))$ according to $\overline{(c)}$.

$\neg(\neg r \vee (p \wedge q))$ [T]		
$\neg r \vee (p \wedge q)$ [F]		
	$(p \wedge q)$ [F]	
$\neg r$ [F]		
r [T]	p [T]	q [F]

We place F next to the subformula $\neg r \vee (p \wedge q)$ on the basis of the assignment of F to its principal subformulas $\neg r$ and $(p \wedge q)$; there is no need to examine the truth-values of the subformulas of $\neg r$ or $(p \wedge q)$. More generally, for every nonatomic $\varphi \in \mathcal{L}$ and every truth-assignment α, $\overline{\alpha}(\varphi)$ depends on just the value(s) of $\overline{\alpha}$ applied to the principal subformula(s) of φ. The semantics of $\mathcal{L}$ are therefore said to be *compositional*, in addition to being truth functional.[7]

We've gone to some trouble to distinguish a truth-assignment α from its extension $\overline{\alpha}$. The former applies only to atomic formulas (variables) whereas

[6] For "principal subformula," see Definition (18) in Section 3.6.

[7] The connectives of natural languages like English often violate truth functionality, as will be discussed in Section 8.4.4, below. In contrast, natural languages are widely believed to be compositional (see [63, Ch. 1]); but there are tantalizing counterexamples [47, p. 108].

the latter applies to all formulas. But let us now dump $\overline{\alpha}$ overboard, and talk henceforth only in terms of α. The following definition makes this possible.

(13) DEFINITION: Let $\varphi \in \mathcal{L}$ and $\alpha \in \mathsf{TrAs}$ be given.

(a) We write $\alpha \models \varphi$ just in case $\overline{\alpha}(\varphi) = \mathsf{T}$, and we write $\alpha \not\models \varphi$ just in case $\overline{\alpha}(\varphi) = \mathsf{F}$.

(b) If $\alpha \models \varphi$ then we say that α ***makes*** φ ***true*** or that α ***satisfies*** φ.

(c) If $\alpha \models \varphi$, we also say that φ ***is true*** according to α. We say that φ ***is false*** according to α if $\alpha \not\models \varphi$.

For example, consider again truth-assignment (d) in Table (3). We saw above that $\overline{(d)}((p \vee q) \rightarrow p) = \mathsf{T}$. Hence, by Definition (13), $(d) \models (p \vee q) \rightarrow p$. We can therefore also say that (d) makes $(p \vee q) \rightarrow p$ true, or that $(p \vee q) \rightarrow p$ is true according to (d).

The new symbol $\models$ is often called "double turnstile." We'll be seeing it often. Unfortunately, this same symbol is used in several different (although related) ways. To not become confused, you have to pay attention to what is on the left and right sides of $\models$ (this tells which use of the symbol is at issue). In Definition (13), we see a truth-assignment α to the left and a formula φ to the right. So we know that $\models$ is being used in the sense (defined above) of satisfaction or "making true."

(14) EXERCISE: (advanced) Let us add a *unary* connective to $\mathcal{L}$, denoting it by $\star$. The new connective is "unary" in the sense that it applies to single formulas φ to make new formulas, $\star\varphi$. Thus, $\star$ works like $\neg$, which is also unary. The language that results from adding $\star$ to $\mathcal{L}$ will be called $\mathcal{L}(\star)$. Its formulas are defined via Definition (8) of Section 3.3 outfitted with the additional clause:

Suppose that φ is a formula. Then so is $\star\varphi$.

For example, $\star r$, $\star\neg q$, and $\neg(\star q \wedge r)$ are formulas of $\mathcal{L}(\star)$. The semantics of $\mathcal{L}(\star)$ are defined by adding the following clause to Definition (6).

> Suppose that φ has the form $\star\psi$, and that $\overline{\alpha}(\chi)$ has already been defined for every subformula χ of ψ. Then $\overline{\alpha}(\varphi) = \mathsf{T}$ if
>
> $$\{\chi \in \mathcal{L}(\star) \mid \chi \text{ is a subformula of } \psi \text{ and } \overline{\alpha}(\chi) = \mathsf{T}\}$$
>
> has an even number of members; otherwise, $\overline{\alpha}(\varphi) = \mathsf{F}$.

Show that $\mathcal{L}(\star)$ is truth functional but not compositional. Then (if you're still with us) devise a semantics for $\mathcal{L}$ that is compositional but not truth functional.

4.2.4 Tables for the connectives

There is an illuminating way to picture Definition (6). The last five clauses of the definition can each be associated with a table that exhibits the truth value of φ as a function of the truth values of φ's principal subformulas. Fact (12)b shows that in computing the truth value of a formula it is enough to consider the truth values of its principal subformulas.

Here is the table for negation.

(15) TABLE FOR NEGATION:

$\neg$	ψ
F	T
T	F

Do you see how the table works? The first line says that if a truth-assignment makes ψ true then it makes $\neg\psi$ false, and vice versa for the second line.

Here are the tables for conjunction and disjunction.

(16) TABLE FOR CONJUNCTION:

χ	$\wedge$	ψ
T	T	T
T	F	F
F	F	T
F	F	F

(17) TABLE FOR DISJUNCTION:

χ	$\vee$	ψ
T	T	T
T	T	F
F	T	T
F	F	F

The second line of (16) says that if a truth-assignment satisfies χ but not ψ, then it makes $\chi \wedge \psi$ false. The second line of (17) says that in the same circumstances the truth-assignment satisfies $\chi \vee \psi$. Here are the two remaining tables.

(18) TABLE FOR CONDITIONALS:

χ	$\rightarrow$	ψ
T	T	T
T	F	F
F	T	T
F	T	F

(19) TABLE FOR BICONDITIONALS:

χ	$\leftrightarrow$	ψ
T	T	T
T	F	F
F	F	T
F	T	F

Observe that Table (18) makes $\chi \rightarrow \psi$ false only if χ is T and ψ is F. In the three remaining cases, $\chi \rightarrow \psi$ is T. [We have already pointed out this feature of Definition (6)e.] Also observe that Table (19) assigns T to $\chi \leftrightarrow \psi$ just in case χ and ψ are assigned the same truth value. Tables (15) - (19) are called *truth tables* for their respective connectives.

To proceed please recall our discussion of "partitions" in Section 2.8. The truth tables for the five connectives rely on partitions of the set TrAs of all truth-assignments. For example, given a particular choice of formula ψ, Table (15) partitions TrAs into two equivalence classes, namely, (i) the truth-assignments that satisfy ψ, and (ii) the truth-assignments that don't satisfy ψ. The table exhibits the truth value of $\neg\psi$ according to the truth-assignments in these two sets. Similarly, Table (16) partitions TrAs into four equivalence classes, namely,

(i) the truth-assignments that satisfy both χ and ψ, (ii) the truth-assignments that satisfy χ but not ψ, (iii) the truth-assignments that satisfy ψ but not χ, and (iv) the truth-assignments that satisfy neither χ nor ψ. The truth value of $\chi \wedge \psi$ according to each of these four kinds of truth-assignments is then exhibited. The other tables may be interpreted similarly.

4.2.5 Truth tables for formulas

Fact (12)a allows us to extend the idea of a truth table to arbitrary formulas. Take the formula $p \rightarrow (q \wedge p)$. We partition the truth-assignments according to what they say about just the variables p, q. The truth-assignments within a given equivalence class of the partition behave the same way on all of the subformulas of $p \rightarrow (q \wedge p)$. So we can devote a single row of the table to each equivalence class of the partition. The columns in a given row are filled with the truth values of the subformulas of $p \rightarrow (q \wedge p)$ according to the truth-assignment belonging to the row. This will be clearer with some examples. The table for $p \rightarrow (q \wedge p)$ is as follows.

(20)

1	2	3	4	5
p	$\rightarrow$	$(q$	$\wedge$	$p)$
T	T	T	T	T
T	F	F	F	T
F	T	T	F	F
F	T	F	F	F

Table (20) is called a *truth table* for $p \rightarrow (q \wedge p)$. To explain it, we'll use the numbers that label the columns in the table. Columns 1 and 3 establish the partition of truth-assignments into four possibilities, namely, those that make both p and q true, those that make p true but q false, etc. Column 5 agrees with column 1 since both record the truth value of p according to the same truth-assignments (otherwise, one of the rows would represent truth-assignments that say that p is both truth and false, and there are no such truth-assignments). Column 4 shows the respective truth values of the subformula $q \wedge p$ according to the four kinds of truth-assignments recorded in the four rows of the table. These truth values are computed using Table (16) on the basis of columns 3 and 5. Column 2 shows the respective truth values of $p \rightarrow (q \wedge p)$. They are computed using Table (18) on the basis of columns 1 and 4. Since column 2 holds the principal connective of the formula, the truth value of $p \rightarrow (q \wedge p)$ appears in this column.[8] Thus, we see that if a truth-assignment satisfies both p and q then it satisfies $p \rightarrow (q \wedge p)$ (this is the first line of the table). If it satisfies p but not q then it makes the formula false (this is the second line of the table. If it fails to satisfy p but does satisfy q then the formula comes out true (third row), and if it satisfies neither p nor q then the formula also comes out true (fourth row). The order of the rows in Table (20) is not important. If we switched the last and second-to-last rows, the table would provide the same information as before.

Here is another truth table, this time without the column numbers, which are inessential.

(21)

$(q$	$\vee$	$p)$	$\wedge$	$\neg$	p
T	T	T	F	F	T
F	T	T	F	F	T
T	T	F	T	T	F
F	F	F	F	T	F

To build Table (21) you first fill in the columns under p and q, being careful to capture all four possibilities and to be consistent about the two occurrences of p. Then you determine the truth values of the subformulas of $(q \vee p) \wedge \neg p$ in the order determined by its construction table.[9] Thus, you determine the truth values of $\neg p$ and place them under the principal connective of this subformula (namely, the tilde). Then you proceed to $q \vee p$, placing truth values under its principle connective, the $\vee$. In truth, you can perform the last two steps in either order, since these two subformulas don't share any occurrences of variables. Finally, we arrive at the subformula $(q \vee p) \wedge \neg p$ itself. Its truth values are placed under the principal connective $\wedge$, which tells us which kinds of truth-assignments satisfy the formula. The first row, for example, shows that truth-assignments that satisfy both p and q fail to satisfy $(q \vee p) \wedge \neg p$.

[8] For "principal connective," see Section 3.6.

[9] For the construction table of a formula see Section 3.3.

Let's do one more example, this time involving three variables hence eight kinds of truth-assignments.

(22)

(	r	$\wedge$	q	)	$\leftrightarrow$	(	p	$\vee$	$\neg$	q	)
	T	T	T		T		T	T	F	T	
	F	F	T		F		T	T	F	T	
	T	F	F		F		T	T	T	F	
	F	F	F		F		T	T	T	F	
	T	T	T		F		F	F	F	T	
	F	F	T		T		F	F	F	T	
	T	F	F		F		F	T	T	F	
	F	F	F		F		F	T	T	F	

From the table's second row we see that any truth-assignment that satisfies p, q, but not r fails to satisfy $(r \wedge q) \leftrightarrow (p \vee \neg q)$.

To get the hang of truth tables, you must construct a few for yourself.

(23) EXERCISE: Write truth tables for:

(a) p

(b) $r \wedge \neg q$

(c) $\neg(q \vee p)$

(d) $(r \wedge \neg q) \rightarrow p$

(e) $p \rightarrow (r \rightarrow q)$

(f) $(p \rightarrow r) \rightarrow q$

(g) $p \wedge \neg(q \leftrightarrow \neg p)$

(h) $(r \wedge q) \leftrightarrow (r \wedge p)$

(i) $(r \wedge r) \vee r$

(j) $(r \wedge \neg q) \rightarrow q$

(k) $(\neg r \wedge q) \rightarrow q$

(l) $p \wedge (q \vee (r \wedge s))$

4.3 Meanings

You sure know a lot about truth-assignments now! It's time to get back to meanings. In this section we'll say what they are (according to Sentential Logic), and discuss some of their properties. The next section explains which meanings are expressed by which formulas.

4.3.1 Truth-assignments as possible worlds

Are truth-assignments meanings? Not quite, but we're only one step away. The step will make sense to you if you share with us a certain conception of truth-assignments.

First, recall that a sentential variable is a logical stand-in for a declarative sentence (say of English) with a determinate truth value. Any such sentence can play this role, including such far-fetched choices as:

(24) (a) The area of a square is the length of its side raised to the power 55.

(b) Zero added to itself is zero.

(c) Bob weighs more than Jack.

(d) Jack weighs more than Bob.

Allowing variables to represent sentences like (24) is compatible with all of the definitions and facts to be presented in the pages that follow. But they are not well-suited for developing the intuitions that underlie much of our theory. More intuitive choices allow each sentence to be true, and allow each to be false. Thus, a nice selection for the sentence represented by a given variable would be "Man walks on Mars before 2050" but not "Triangles have three sides." The former might be true, and it might be false (we can't tell right now), but the latter can only be true.[10] Intuitions may get messed up if you allow a sentential variable to stand for a sentence that is necessarily true (or necessarily false), as

[10]You have the perfect right to ask us how we know that four-sided triangles are impossible. Who said this is so? Maybe there is a four-sided triangle somewhere in Gary, Indiana. In response to such interrogation, all we can (feebly) respond is that it *seems* to us that, somehow,

in (24)a,b. Intuitions can also get messed up if there are constraints on which variables can be true or false *together*, as in (24)c,d. Instead, the sentences should be logically independent of each other. For example, the three sentences

p : The temperature falls below freezing in New York on Labor Day 2010.
q : The temperature falls below freezing in Chicago on Labor Day 2010.
r : The temperature falls below freezing in Minneapolis on Labor Day 2010.

are logically independent; they can be true or false in any combination.

When we are explaining things intuitively, it will henceforth be assumed that you've chosen interpretations of variables that are logically independent. This will spare us from contemplating (bizarre) truth assignments that assign T to both (24)c,d, or F to (24)b. But formally speaking, you're on your own. Variables can represent any declarative sentences that make you happy. (We will avert our gaze.)

If logical independence holds, then the variables stand for sentences that can be true or false in any combination. Each such combination is a "way the world could be" (as they say in famous parlance). It is a possible state of reality (in another idiom). For example, in one possible state of reality p is true, q is false, and r is false. If the variables are as above then in this possible state of reality New York freezes on Labor Day 2010 but Chicago and Minneapolis escape the cold weather. In another possible world, all three variables are true. In brief, each combination of truth values for the variables is a *possible world*. Now notice that truth-assignments are nothing but combinations of truth values for the variables. For example, the combination in which p is true, and q, r are false is truth-assignment (d) in Table (3). Each truth-assignment can thus be conceived as a possible world. Since there are eight truth-assignments over three variables, there are eight possible worlds involving three variables.

It would be more accurate to qualify a truth-assignment as "a possible world

the meaning of the word "triangle" makes four-sidedness an impossibility. We're just trusting you to see things the same way. Anyway, we'll never get on with our business if we try to sort out the issue of geometrical certainty in this book. For that, you'll do better reading Soames [92].

insofar as worlds can be described using the variables of $\mathcal{L}$." For, there is bound to be more to a potential reality than can be expressed using a measly n sentential variables. But from our vantage inside $\mathcal{L}$, all that can be seen is a world's impact on the sentential variables. So, we identify a world with the particular truth-assignment that it gives rise to.

One of the eight truth-assignments is the *true* one, of course. By the true truth-assignment, we mean the one whose truth values correspond to reality — the world as it (really) is. For the meteorological variables in our example, the (real) world assigns true to p if New York is freezing on Labor Day in 2010 and false otherwise, and likewise for q and r. We don't know at present which of the eight truth-assignments is the true one, but that doesn't matter to our point about truth. We claim that one (and only one) of the truth-assignments is true, whether or not we know which one has this virtue.

And we're saying that one of the truth-assignments is true *right now*, not that it will become true on Labor Day in 2010. This idea might be hard to swallow. It is tempting to think that the choice of true truth-assignment is left in abeyance until the weather sorts itself out on the fateful day. But this is not the way we wish to look at the matter. Our point of view will rather be that the future is already a fact, just unknown to us. Variables bearing on the future thus have an (unknown) truth value. Hence, one of the truth-assignments is the true one.

If our variables involved past or present events (e.g., whether Julius Caesar ever visited Sicily) then the idea that exactly one of the truth-assignments is true would be easier to accept. We extend the same idea to future events in order to render our interpretation of Sentential Logic as general as possible — almost any determinate declarative sentences can interpret the variables. Let us admit to you, however, that our breezy talk of "the true truth-assignment" among the set of potential ones is not to everyone's liking. But it is the way we'll proceed in this book.

4.3.2 Meaning as a set of truth-assignments

Since truth-assignments can be conceived as possible worlds, a set of truth-assignments can be conceived as a set of possible worlds; it is a set of ways the world might be. If you could somehow declare this set, you would be declaring "these are the ways the world might be." For example, if you declared the set consisting of truth-assignments $(a), (b), (c), (d)$ from Table (3), you would be asserting that the facts bear out one of these four truth-assignments. Now notice that p is true in each of $(a), (b), (c), (d)$, and in no other truth-assignment.[11] Thus, declaring $\{(a), (b), (c), (d)\}$ amounts to declaring that p is true! As we said above, by p being true, we mean *true in the actual world*, as it really is. Thus, if you assert $\{(a), (b), (c), (d)\}$, you are asserting that p is true in this sense. The latter set is therefore an appropriate interpretation of the "meaning" of the formula p. (What else could you have meant by asserting p, other than that p is true?) In contrast, each of the four combinations of truth and falsity for q and r are realized by one of $(a), (b), (c), (d)$. Therefore, in "declaring" $\{(a), (b), (c), (d)\}$, nothing follows about q and r. Both might be true [as in (a)], both could be false [as in (d)], or just one could be true [as in (b), (c)]. Declaring $\{(a), (b), (c), (d)\}$ does not amount to declaring q, nor to declaring r.

We shall now proceed to generalize the foregoing idea by taking meanings to be arbitrary sets of truth-assignments. In the following definition, remember that the number of sentential variables in $\mathcal{L}$ has been fixed at n.

(25) DEFINITION:

- (a) Reminder [from Definition (5)]: The set of all truth-assignments is denoted TrAs.
- (b) Any subset of TrAs is a *meaning*.
- (c) The set of all meanings is denoted $\mathsf{Meanings}$.

It follows from the definition that $\mathsf{Meanings}$ is a set of sets (analogously to the set $\{\{2, 4\}, \{4, 3\}, \{9, 2\}\}$ of sets of numbers). To illustrate, suppose (as usual)

[11] You're unlikely to simply remember this fact. So you really ought to go back to the table and look.

that there are three variables. Then the set

$$M = \{(a), (b), (c), (e), (f), (g)\}$$

is one member of Meanings, just as $\{(a), (b), (c), (d)\}$ (mentioned above) is another member. Which truth-assignments are missing from M above? Just the two truth-assignments that make both q and r false are missing. The set M thus expresses the assertion that at least one of q, r is true. We'll see that within Sentential Logic this meaning is expressed by the formula $q \vee r$.

How many meanings are there? Well, how may truth-assignments are there? There are 2^n truth-assignments [see Fact (4)]. Each truth-assignment may appear or fail to appear in an arbitrary meaning. To compose a meaning therefore requires 2^n binary choices. Since these choices are independent of each other, all 2^n of them give rise to 2^{2^n} combinations, hence 2^{2^n} meanings. Let us record this fact.

(26) FACT: If n is the number of sentential variables in $\mathcal{L}$ then there are 2^{2^n} meanings. That is, Meanings has 2^{2^n} members.

Thus, with 3 variables there are $2^{2^3} = 2^8 = 256$ meanings. As the number of variables goes up, the number of meanings grows quickly. With just 4 variables, Meanings has $65,536$ members. With 5 variables, there are more than 4 billion meanings. With 10 variables, Meanings is astronomical in size. (Applications of Sentential Logic to industrial settings often involve hundreds of variables; see [30, 42].)

4.3.3 Varieties of meaning

Now you know what meanings are in Sentential Logic. And you were already acquainted with the set of formulas (from Chapter 3). So we're ready to face the pivotal question: Which meaning is expressed by a given formula of $\mathcal{L}$? We defer this discussion for one more moment. There are still some observations to make about meanings themselves.

One subset of TrAs is TrAs itself. Hence, TrAs $\in$ Meanings. But what on earth does TrAs mean? It seems to represent no more than the idea that some truth-assignment gives the truth values of the sentential variables in the real world. We already knew that one of the truth-assignments accomplishes this feat; after all, that's how we set things up (by limiting attention to sentential variables that were either true or false and not both). So TrAs seems to be vacuous as a meaning. Since it doesn't eliminate any possibilities, it doesn't provide any information. Let us not shrink from this conclusion. TrAs is indeed the vacuous meaning, providing no information. We'll have great use for this special case. For example, it will be assigned as meaning to the formulas $p \to p$ and $q \vee \neg q$, among others.

Let's use the symbol Reality to denote the one truth-assignment whose truth values are given by reality (by the "real" world). This notation allows us to define when a given meaning M is true. Since M is the idea that the world conforms to one of its members, M is true just in case Reality belongs to it. Again: $M \in$ Meanings is true if and only if Reality $\in M$. For example, if $M = \{(a), (b), (c), (d)\}$ [relying again on Table (3)], then M is true if Reality $\in M$, that is, if Reality is one of (a), (b), (c), (d).

Please think again about our vacuous meaning TrAs. Is it the case that Reality $\in$ TrAs? Sure. TrAs holds all the truth-assignments, so it must hold Reality, the truth-assignment made true in the real world. But to reach this conclusion, we don't need to know the slightest thing about Reality. We don't need to know what sentences the variables represent, nor whether any particular one of them is true or false. In this sense, the meaning TrAs is guaranteed to be true. That's what makes it vacuous. If you assert something that is guaranteed to be true, no matter what the facts are, then you haven't made any substantive claim at all.

The other limiting case of a meaning is the empty set, $\emptyset$, the set with no members. [It counts as a genuine subset of TrAs; see Fact (13) in Section 2.6.] Can you figure out what $\emptyset$ means? Don't say that it means nothing. Since $\emptyset \subseteq$ TrAs, $\emptyset \in$ Meanings. Our question is: which meaning does $\emptyset$ express?

Roughly, $\emptyset$ is the idea that the actual world is not among the possibilities, in other words, that what's true is impossible. This is evidently false, which accords with the fact that $\mathsf{Reality} \notin \emptyset$ (since nothing is a member of $\emptyset$). Again, we need know nothing about Reality to reach the conclusion that it is not a member of $\emptyset$. So, just as TrAs is guaranteed to be true, $\emptyset$ is guaranteed to be false. We'll need this strange case when we get around to assigning meanings to formulas like $p \wedge \neg p$ and $r \leftrightarrow \neg r$.

In between TrAs and $\emptyset$ lie the meanings that are neither trivially true nor trivially false. Such meanings are called *contingent*.[12] Whether a contingent meaning is true or false depends on what Reality is like. If Reality makes p, q and r all true then the contingent meaning $\{(a), (b)\}$ is true; if Reality falsifies p then this meaning is false.

Now take two contingent meanings M_1 and M_2, and suppose that $M_1 \subset M_2$.[13] Then M_1 makes a stronger claim than M_2 since M_1 situates the (real) world in a narrower class of possibilities than does M_2. We illustrate again with Table (3). The meaning $\{(b), (c)\}$ is the idea that $\mathsf{Reality} \in \{(b), (c)\}$, in other words, that reality conforms to one of $(b), (c)$. This meaning has more content than the idea $\{(b), (c), (e)\}$, which says that $\mathsf{Reality} \in \{(b), (c), (e)\}$, in other words, that reality conforms to one of $(b), (c), (e)$. Among the nonempty meanings, you can see the strongest consist of just one truth-assignment, like the meaning $\{(d)\}$. Such a meaning pins the world down to a single truth-assignment. It specifies that p has truth value so-and-so, q has truth value thus-and-such, and so forth for all the variables. At the opposite side of the spectrum, the "weakest" nontrivial meaning is missing just a single truth-assignment. An example using Table (3) is $\{(b), (c), (d), (e), (f), (g), (h)\}$. This meaning excludes only the possibility that all three sentential variables are true.

Not every pair of meanings can be compared in strength. Neither $\{(b), (c)\}$ nor $\{(c), (d), (e)\}$ is a subset of the other so neither is stronger in the sense we have been discussing. Of course, they might be comparable in some other sense of "strength." Perhaps truth-assignment (b) is more surprising than any of (c), (d), (e). That might suffice for $\{(b), (c)\}$ to be considered a stronger claim than $\{(c), (d), (e)\}$. In Chapter 9 we'll develop an apparatus to clarify this idea.

---

[12]Merriam-Webster offers the following definition of "contingent:" dependent on or conditioned by something else. The "something else" at work in the present context is Reality.

[13]To remind you, this notation means that M_1 is a proper subset of M_2. See Section 2.2.

For the moment, we'll rest content with the subset-criterion of strength, even though it does not allow us to compare every pair of meanings. According to this criterion, the contradictory meaning $\emptyset$ is the strongest since for every other meaning M, $\emptyset \subset M$. Likewise, the vacuous meaning TrAs is weakest since for all other meanings M, $M \subseteq \mathsf{TrAs}$.

(27) EXERCISE: Suppose that there are three variables, and let the eight truth-assignments be as shown in Table (3). Indicate some pairs of meanings in the following list that are comparable in content, and say which member of each such pair is stronger.

(a) $\{(b), (c)\}$

(b) $\{(a), (c)\}$

(c) $\{(b), (c), (h)\}$

(d) $\{(a), (c), (f), (h)\}$

(e) $\{(a), (b), (c), (d), (e), (f), (g), (h)\}$

(f) $\{(a), (c), (d), (e), (f), (g), (h)\}$

(g) $\emptyset$

(h) $\{(c), (d), (e), (f), (g), (h)\}$

(i) $\{(c)\}$

(j) $\{(b)\}$

4.4 Meanings of formulas

Prepare yourself. The time has come to attach meanings to formulas. Specifically, for every formula $\varphi \in \mathcal{L}$, we now define its meaning. As discussed above, the meaning of φ will be a set of truth-assignments. To denote the latter set, we use the notation $[\varphi]$.

4.4.1 The key definition

What meaning should be assigned to the sentential variable p? That is, what should we take as $[p]$? This case was discussed in Section 4.3.2, above. If $n = 3$, then $\{(a),(b),(c),(d)\}$ holds all and only the truth-assignments in which p is true. This set embodies the idea that p is true, hence the set constitutes its meaning. Generalizing, we see that it makes sense to assign as meaning to a sentential variable v_i the set of all truth-assignments in which v_i is true. That is, $[v_i] = \{a \in \mathsf{TrAs} \mid a(v_i) = \mathsf{T}\}$.[14] So now you know what meaning is attached to atomic formulas (that is, to sentential variables). But what about nonatomic formulas? What meaning do they get? We think you've guessed the answer already. It is given in the following definition.

(28) DEFINITION: Let formula $\varphi \in \mathcal{L}$ be given. Define:

$$[\varphi] = \{\alpha \in \mathsf{TrAs} \mid \alpha \models \varphi\}.$$

We call $[\varphi]$ *the meaning of* φ.

Thus, the meaning of φ is the set of truth-assignments that satisfy it. Definition (28) just extends our understanding of the meaning of atomic formulas to all formulas, relying for this purpose on Definitions (6) and (13).

To illustrate, suppose again that $\mathcal{L}$ has just three variables, and consider the formula $p \rightarrow (q \wedge p)$ From its truth table (20) and Table (3), we see that $p \rightarrow (q \wedge p)$ is satisfied by truth-assignments $(a),(b),(e),(f),(g),(h)$. Hence,

$$[p \rightarrow (q \wedge p)] = \{(a),(b),(e),(f),(g),(h)\}.$$

Consider now formula $(q \vee p) \wedge \neg p$. From its truth table (21), we see that just $(e),(f)$ satisfy it. Hence,

$$[(q \vee p) \wedge \neg p] = \{(e),(f)\}.$$

[14]The last equation may be read as follows. The meaning of the sentential variable v_i is the set of truth assignments that map v_i to the truth value T.

Finally, consider $(r \wedge q) \leftrightarrow (p \vee \neg q)$. Its truth table (22) shows that:

$$[(r \wedge q) \leftrightarrow (p \vee \neg q)] = \{(a), (f)\}.$$

Let us follow up a remark made in Section 4.3.3, above. Recall that the set of all meanings is denoted by Meanings. Each member of Meanings is a subset of TrAs, the set of all truth-assignments. We noted earlier that $M \in$ Meanings is true if and only if Reality $\in M$, where Reality is the truth-assignment that corresponds to the facts. The same observation extends to formulas. Given a formula φ, $[\varphi]$ is a subset of TrAs. So, we say that a formula is *true* just in case Reality $\in [\varphi]$. By Definition (28), Reality $\in [\varphi]$ holds just in case Reality $\models \varphi$. In words, a formula is true just in case its meaning includes reality, that is, just in case reality makes it true. Doesn't this make perfect sense?

(29) EXERCISE: Suppose that there are three variables, and let the eight truth-assignments be as shown in Table (3). Compute the meanings of the following formulas. (That is, write down the truth-assignments that fall into each meaning. Use the notation $[\varphi]$.)

(a) q

(b) $\neg q \wedge p$

(c) $\neg q \rightarrow r$

(d) $r \wedge (q \vee r)$

(e) $r \rightarrow (p \vee r)$

(f) $r \leftrightarrow (p \wedge q)$

(g) $q \wedge (q \rightarrow \neg q)$

4.4.2 Meanings and set operations

The examples of Section 4.4.1 show that one way to compute the meaning of a nonatomic formula is via its truth table. The members of $[\varphi]$ are the truth-assignments that yield T under the principal connective in the truth table for φ. You might think of this as the "bottom up" approach to calculating $[\varphi]$. There is also a "top down" perspective that is worth understanding. It is embodied in the following fact.

(30) FACT: Let nonatomic formula $\varphi \in \mathcal{L}$ be given.[15]

(a) Suppose that φ is the negation $\neg\psi$. Then $[\varphi] = \mathsf{TrAs} - [\psi]$.

(b) Suppose that φ is the conjunction $\chi \wedge \psi$. Then $[\varphi] = [\chi] \cap [\psi]$.

(c) Suppose that φ is the disjunction $\chi \vee \psi$. Then $[\varphi] = [\chi] \cup [\psi]$.

(d) Suppose that φ is the conditional $\chi \rightarrow \psi$. Then $[\varphi] = (\mathsf{TrAs} - [\chi]) \cup [\psi]$.

(e) Suppose that φ is the biconditional $\chi \leftrightarrow \psi$. Then $[\varphi] = ([\chi] \cap [\psi]) \cup ((\mathsf{TrAs} - [\chi]) \cap (\mathsf{TrAs} - [\psi]))$.

We can summarize the fact as follows.

(31) $[\neg\psi] = \mathsf{TrAs} - [\psi]$

$[\chi \wedge \psi] = [\chi] \cap [\psi]$.

$[\chi \vee \psi] = [\chi] \cup [\psi]$.

$[\chi \rightarrow \psi] = (\mathsf{TrAs} - [\chi]) \cup [\psi]$.

$[\chi \leftrightarrow \psi] = ([\chi] \cap [\psi]) \cup ((\mathsf{TrAs} - [\chi] \cap (\mathsf{TrAs} - [\psi]))$.

The five clauses of Fact (30) follow directly from Definitions (28) and (6). Consider (30)b, for example. A given truth-assignment α belongs to $[\chi \wedge \psi]$ just in case $\alpha \models \chi \wedge \psi$ [this is what Definition (28) says]. And according to Definition (6)c, $\alpha \models \chi \wedge \psi$ just in case $\alpha \models \chi$ and $\alpha \models \psi$. By Definition (28) again, $\alpha \models \chi$ and $\alpha \models \psi$ if and only if $\alpha \in [\chi]$ and $\alpha \in [\psi]$, which is true if and only if $\alpha \in [\chi] \cap [\psi]$. So we've shown that $\alpha \in [\chi \wedge \psi]$ if and only if $\alpha \in [\chi] \cap [\psi]$. This proves (30)b. The other clauses are demonstrated similarly.

[15]Reminder: $\mathsf{TrAs} - [\psi]$ is the set of truth-assignments that do not belong to $[\psi]$. See Section 2.3. $[\chi] \cap [\psi]$ is the set of truth-assignments that belong to both $[\chi]$ and $[\psi]$. See Section 2.4. $[\chi] \cup [\psi]$ is the set of truth-assignments that belongs to either or both $[\chi]$ and $[\psi]$. See Section 2.5.

Let's use (30)a to compute $[\neg r]$. From Table (3):

$$[r] = \{(a), (c), (e), (g)\}.$$

According to (30)a, $[\neg r]$ is the complement of TrAs. Hence:

$$[\neg r] = \{(b), (d), (f), (h)\}.$$

Isn't this outcome reasonable assuming that $\neg$ corresponds (roughly) to "not" in English? The set $\{(b), (d), (f), (h)\}$ contains exactly the truth-assignments in which r is false. More generally, for an arbitrary formula ψ, the complement of $[\psi]$ (relative to TrAs) contains exactly the truth-assignments in which ψ is false. This is the sense in which $\neg$ translates English negation, notably, "not."

Next, let's compute $[p \wedge q]$. Table (3) informs us that $[p] = \{(a), (b), (c), (d)\}$ and $[q] = \{(a), (b), (e), (f)\}$. According to (30)b, $[p \wedge q]$ is the intersection of the latter sets, hence $[p \wedge q] = \{(a), (b), (c), (d)\} \cap \{(a), (b), (e), (f)\} = \{(a), (b)\}$. Again, this outcome is reasonable assuming that the wedge corresponds (roughly) to "and" in English. (See Section 3.7.) The set $\{(a), (b)\}$ contains exactly the truth-assignments in which both p and q are true. More generally, for arbitrary formulas χ, ψ, the intersection of $[\chi]$ with $[\psi]$ contains exactly the truth-assignments in which both χ and ψ are true. This is the sense in which $\wedge$ translates English conjunction "and."

For a more complicated illustration, let us compute $[\neg r \vee (p \wedge q)]$ according to Fact (30). The examples just reviewed show that $[\neg r] = \{(b), (d), (f), (h)\}$ and $[p \wedge q] = \{(a), (b)\}$. According to (30)c, $[\neg r \vee (p \wedge q)]$ is the union of the latter sets, hence $[\neg r \vee (p \wedge q)] = \{(b), (d), (f), (h)\} \cup \{(a), (b)\} = \{(a), (b), (d), (f), (h)\}$. If you look through the truth-assignments listed in Table (3), you'll see that each member of $\{(a), (b), (d), (f), (h)\}$ either makes r false or makes both p, q true, or does both of these things. None of the other truth-assignments have this property. So, $\{(a), (b), (d), (f), (h)\}$ is the appropriate meaning for the idea that either $\neg r$ or $p \wedge q$ is true (or both). More generally, for arbitrary formulas χ, ψ, the union of $[\chi]$ with $[\psi]$ contains exactly the truth-assignments in which either or both χ of ψ are true. This is the sense in which $\vee$ translates English disjunction "or."

Of course, in all these cases we obtain the same meaning using set operations as we do using a truth table. This is guaranteed by Fact (30).

The meaning assigned to conditionals is not as intuitive as the other connectives. Consider the meaning assigned to $p \rightarrow q$ by (30)d. We know that $[p] = \{(a), (b), (c), (d)\}$ and $[q] = \{(a), (b), (e), (f)\}$. Hence, $\mathsf{TrAs} - [p] = \{(e), (f), (g), (h)\}$. Clause (30)d thus dictates that $[p \rightarrow q] = \{(e), (f), (g), (h)\} \cup \{(a), (b), (e), (f)\} = \{(a), (b), (e), (f), (g), (h)\}$. The latter set includes all truth-assignments except (c) and (d). What do you notice about these two truth-assignments? [Hint: To answer this question, you need to look at Table (3)!] Yes, (c) and (d) have the particularity of declaring p to be true and q to be false. These are the only truth-assignments missing from $[p \rightarrow q]$. So, $[p \rightarrow q]$ seems to embody the idea that reality can be anything that doesn't make p true and q false. More generally, $[\varphi \rightarrow \psi]$ holds every truth-assignment except for those that satisfy φ but not ψ. We already commented on this feature of conditionals in Section 4.2.2.

It is left to you to compare the "bottom up" and "top down" approaches to the meaning of biconditionals.

(32) EXERCISE: Suppose $\mathcal{L}$ contains just three variables, and let the truth-assignments be listed as in (3). Use set operations to calculate the meanings of the following formulas. Proceed step by step, as we did to illustrate Fact (30).

(a) $q \vee \neg r$

(b) $(p \vee \neg r) \wedge q$

(c) $p \rightarrow (q \vee \neg r)$

(d) $p \leftrightarrow (q \vee \neg r)$

(e) $p \leftrightarrow (q \vee \neg p)$

(f) $p \rightarrow (p \vee \neg r)$

4.4.3 Long conjunctions and long disjunctions

In Section 3.8 we noted that it is often convenient to abbreviate formulas like $((p \wedge q) \wedge r)$ to just $p \wedge q \wedge r$. Similarly, we like to write $((p \rightarrow q) \wedge q) \wedge (r \vee q)$ as

$(p \to q) \wedge q \wedge (r \vee q)$, just as we like to write $(p \to q) \vee q \vee (r \vee q)$ in place of either $((p \to q) \vee q) \vee (r \vee q)$ or $(p \to q) \vee (q \vee (r \vee q))$.

What allows us to drop parentheses in these cases is the identical meanings of the formulas $\varphi \wedge (\psi \wedge \chi)$ and $(\varphi \wedge \psi) \wedge \chi$ — and likewise for disjunctions. The matter is expressed the following fact, whose truth should be clear to you by now.

(33) FACT: Let formulas φ, ψ, χ be given. Then:

(a) $[(\varphi \wedge \psi) \wedge \chi] = [\varphi \wedge (\psi \wedge \chi)]$.

(b) $[(\varphi \vee \psi) \vee \chi] = [\varphi \vee (\psi \vee \chi)]$.

Similar equalities hold for more than three formulas.

4.5 A look forward

Now you know what meanings are in Sentential Logic, and which formulas express which of them. So in addition to Spanish or French (or whatever else you speak), you have become fluent in $\mathcal{L}$. This is quite an achievement. Congratulations! But there is more work ahead as we attempt to put $\mathcal{L}$ to use. Recall from Section 1.1 that logic is meant to be an aid to thought, protecting us from missteps leading to fallacy. To serve this purpose, $\mathcal{L}$ must be brought to bear on reasoning, and an attempt made to distinguish secure patterns of inference from less secure. The next chapter initiates precisely this task. It introduces the idea of a "valid" argument as the formal counterpart of a secure inference. It also identifies some formulas as "logical truths," which means that their truths are secure without introducing any assumptions at all. Doesn't this sound interesting?

Take a short break. We'll see you in Chapter 5.

Chapter 5

Validity and other semantic concepts

Now you know what meanings are in Sentential Logic, right? (If not, you'll have to go back and plow through Chapter 4 once again.) In the present chapter we reap the harvest of our hard work. Through the idea of "the meaning of a formula" we shall explain fundamental ideas of Logic including *validity, tautology*, and *contradiction*. You'll see that the three ideas are closely related to each other. Let's start with validity.

5.1 Validity

5.1.1 Arguments

The word "argument" in English connotes disputation but logicians drain the term of its bellicose overtones. Just the list of statements offered by a given side in the debate is taken into account. Formally, we proceed as follows.

(1) DEFINITION: By an *argument (of $\mathcal{L}$)* is meant a nonempty, finite list $\varphi_1 \dots \varphi_k, \psi$ of formulas. The last formula on the list, ψ, is called the argument's *conclusion*. The remaining formulas, $\varphi_1 \dots \varphi_k$, are called the argument's *premises*.

To demarcate the conclusion from the premises, we often replace the last comma with a slash. Thus, $p \rightarrow q, \neg(p \vee r)\ /\ \neg q$ is an argument with conclusion $\neg q$ and premises $p \rightarrow q$ and $\neg(p \vee r)$. When we consider arguments in English (instead of $\mathcal{L}$), it will be convenient to write them as vertical lists with the conclusion separated from the premises by a horizontal line. You saw examples in Section 1.4. We observe for future reference that in Definition (1), k might be zero. In other words, an argument might have no premises at all (but it always has a conclusion).

This brings us to a fundamental question for Logic. Which arguments represent good reasoning, and which represent bad?

5.1.2 Validity as inclusion

Sentential Logic makes a sharp distinction between good and bad arguments. (Good arguments will be called "valid.") In Logic, an argument is good just in case its premises *force* its conclusion to be true. You may wonder how such compulsion between premises and conclusion could arise. The following notation will help us understand the matter.

(2) DEFINITION: Given a set Γ of formulas, we denote by $[\Gamma]$ the collection of truth-assignments that satisfy every member of Γ.[1] In other words, if $\Gamma = \{\varphi_1 \dots \varphi_k\}$, then

$$[\Gamma] = \{\alpha \in \mathsf{TrAs} \,|\, \alpha \models \varphi_1 \text{ and } \dots \text{ and } \alpha \models \varphi_k\}.$$

In the definition, the expression $\{\alpha \in \mathsf{TrAs} \,|\, \alpha \models \varphi_1 \text{ and } \dots \text{ and } \alpha \models \varphi_k\}$ represents the set of all truth-assignments α having the property that α satisfies each of φ_1 through φ_k. For an example, suppose again that $\mathcal{L}$ has three variables, and consider the eight truth-assignments named in Table (3) of Section 4.2.1, repeated here for convenience.

(3)

	p	q	r
(a)	T	T	T
(b)	T	T	F
(c)	T	F	T
(d)	T	F	F
(e)	F	T	T
(f)	F	T	F
(g)	F	F	T
(h)	F	F	F

Then $[p, r] = \{(a), (c)\}$ and $[q \rightarrow r, \neg r] = \{(d), (h)\}$. Given the way conjunction is interpreted in Sentential Logic [see (6)c of Section 4.2.2], the following fact should be transparent.

[1]The Greek letter Γ is pronounced: "(capital) gamma".

(4) FACT: Let $\Gamma = \{\varphi_1 \dots \varphi_k\}$ be a set of formulas. Then $[\Gamma] = [\varphi_1 \wedge \dots \wedge \varphi_k]$.

For example $[p, q \rightarrow r, \neg r] = [p \wedge (q \rightarrow r) \wedge \neg r] = \{(d)\}$.

The point about $[\Gamma]$ is that it is a meaning, embodying the assertion of all the members of Γ. Note the use of the brackets in $[\Gamma]$, which were introduced in Definition (28) of Section 4.4.1 to denote the meaning of formulas. Definition (2) above simply extends this notation to sets of formulas.

If you assert all of $\varphi_1 \dots \varphi_k$ then you are claiming that Reality satisfies each of the φ_i, hence that it satisfies their conjunction, hence that it falls in $[\varphi_1 \dots \varphi_k]$.[2] For example, if you assert both $q \rightarrow r$, and $\neg r$, then you are claiming that Reality is one of $\{(d), (h)\}$. If you assert all of p, $q \rightarrow r$, $\neg r$ then you claim that Reality is the truth-assignment (d).

Now consider an argument $\varphi_1 \dots \varphi_k / \psi$. Suppose that $[\varphi_1 \dots \varphi_k]$ is a subset of $[\psi]$. That is, suppose that every truth-assignment satisfying all of the premises also satisfies ψ. Then if the premises are true, the conclusion must be true as well. For, the truth of the premises amounts to the claim that Reality is a member of $[\varphi_1 \dots \varphi_k]$; and since $[\psi]$ includes $[\varphi_1 \dots \varphi_k]$, Reality must also be a member of $[\psi]$. But as we observed in Section 4.4.1, to say that Reality belongs to $[\psi]$ is just to say that ψ is true. The upshot is that if $[\varphi_1 \dots \varphi_k] \subseteq [\psi]$ then the truth of $\varphi_1 \dots \varphi_k$ guarantees the truth of ψ. We are led by this reasoning to the following definition.

(5) DEFINITION: Let argument $\varphi_1 \dots \varphi_k / \psi$ be given.

   (a) $\varphi_1 \dots \varphi_k / \psi$ is *valid* just in case $[\varphi_1 \dots \varphi_k] \subseteq [\psi]$. Otherwise, it is *invalid*.

   (b) If $\varphi_1 \dots \varphi_k / \psi$ is valid we write $\{\varphi_1 \dots \varphi_k\} \models \psi$. If it is invalid we write $\{\varphi_1 \dots \varphi_k\} \not\models \psi$.

   (c) If $\{\varphi_1 \dots \varphi_k\} \models \psi$ then we also say that:

      i. $\{\varphi_1 \dots \varphi_k\}$ *implies* ψ,

[2] Reminder: Reality denotes the truth-assignment that maps each variable to the truth value it enjoys in the real world. See Section 4.3.1. By the way, the formula φ_i is a standard way of denoting an arbitrary member of $\varphi_1 \dots \varphi_k$.

ii. ψ *follows (logically) from* $\{\varphi_1 \ldots \varphi_k\}$, and
iii. ψ *is a (logical) consequence of* $\{\varphi_1 \ldots \varphi_k\}$.

For example, $[p \wedge q] = \{(a), (b)\}$ whereas $[q] = \{(a), (b), (e), (f)\}$. Hence $[p \wedge q] \subseteq [q]$, so $p \wedge q/q$ is a valid argument. Equivalently, we say that $\{p \wedge q\}$ implies q and write $\{p \wedge q\} \models q$. To reduce clutter, we often drop the brackets on the left side of $\models$. Thus, the validity of $p \wedge q/q$ may be written as $p \wedge q \models q$, and we say that $p \wedge q$ implies q. For another example, consider the argument $(p \vee q) \leftrightarrow r, q/r$. You can compute that $[(p \vee q) \leftrightarrow r] = \{(a), (c), (e), (h)\}$ and $[q] = \{(a), (b), (e), (f)\}$, so $[(p \vee q) \leftrightarrow r, q] = \{(a), (e)\}$. Moreover, $[r] = \{(a), (c), (e), (g)\}$ Hence, $[(p \vee q) \leftrightarrow r, q] \subseteq [r]$, so $(p \vee q) \leftrightarrow r, q/r$ is valid and we write $(p \vee q) \leftrightarrow r, q \models r$. That is, $(p \vee q) \leftrightarrow r, q$ implies r. On the other hand, $[(p \wedge q) \leftrightarrow r] = \{(a), (d), (f), (h)\}$ so $[(p \wedge q) \leftrightarrow r, q] = \{(a), (f)\}$. Thus $[(p \wedge q) \leftrightarrow r, q] \not\subseteq [r]$, so $(p \wedge q) \leftrightarrow r, q/r$ is invalid and we write $(p \wedge q) \leftrightarrow r, q \not\models r$, and say that $(p \wedge q) \leftrightarrow r, q$ does not imply r.

To affirm that a whole class of arguments is valid, we sometimes revert to Greek. Thus, we write $\varphi \wedge \psi \models \varphi$ to affirm:

$$p \wedge q \models p, \quad (r \vee t) \wedge (p \rightarrow q) \models (r \vee t), \quad \neg(r \rightarrow q) \wedge q \models \neg(r \rightarrow q),$$

and so forth. Using such notation allows us to state two familiar principles of reasoning, along with their Latin names. (They will figure in the developments of Chapter 10). The first is illustrated in English by the inference: If Windows is defective then Microsoft will ultimately go broke. Windows is defective. Therefore Microsoft will ultimately go broke. The second is illustrated by: If Windows is defective then Microsoft will ultimately go broke. Microsoft will never go broke. Therefore, Windows is not defective.

(6) FACT:

(a) MODUS PONENS: $\{\varphi \rightarrow \psi, \varphi\} \models \psi$
(b) MODUS TOLLENS: $\{\varphi \rightarrow \psi, \neg\psi\} \models \neg\varphi$

We prove the fact. Regarding Modus Ponens, consider a truth-assignment α that makes the conclusion false. We'll show that α makes at least one premise false. Either α makes φ true or it makes φ false. If it makes φ false then it makes the second premise false; if it makes φ true then it makes the first premise false (because by hypothesis α makes ψ false). Consequently, there is no truth-assignment that makes the conclusion false and the premises both true. In other words, every truth-assignment that makes both premises true also makes the conclusion true. So, by Definition (5), the premises imply the conclusion.

Modus Tollens is established similarly. Consider a truth-assignment α that makes the conclusion false. Then α makes φ true. We'll show that α makes at least one premise false. Either α makes ψ true or it makes ψ false. If it makes ψ true then it makes the second premise false; if it makes ψ false then it makes the first premise false (because by hypothesis α makes φ true). Consequently, there is no truth-assignment that makes the conclusion false and the premises both true. Definition (5) may thus be invoked, as before.

Did you notice that Definition (5) entrusts $\models$ with a second mission? Definition (13) of Section 4.2.2 has $\models$ relating truth-assignments to formulas. For example, we write $(a) \models p \vee \neg q$ to denote the fact that $p \vee \neg q$ is true in the truth-assignment (a). Starting with Definition (5), we also use $\models$ to relate sets of formulas to another formula. For example, we write $\{p \wedge q\} \models q$ (or more succinctly, $p \wedge q \models q$) to signify that $[p \wedge q] \subseteq [q]$. Yes, doubling up the use of $\models$ invites confusion. But there is nothing to be done about it; generations of logicians write $\models$ in both senses. Just remember that when we write $\alpha \models \varphi$, we're talking about the satisfaction relation between a truth-assignment and a formula. When we write $\{\varphi_i \ldots \varphi_k\} \models \psi$, we're talking about a relation of inclusion between the meaning of $\varphi_i \wedge \ldots \wedge \varphi_k$ and the meaning of ψ. The thing to the left of $\models$ tells you which interpretation of $\models$ is at issue.

Please be careful about the status of $\models$. It is not a symbol of $\mathcal{L}$. ($\mathcal{L}$ was entirely specified in Chapter 3, where $\models$ is not mentioned.) Rather, $\models$ is just an extension of English that allows us to concisely express facts about satisfaction and validity. So you must not write, for example, $r \rightarrow ((p \wedge q) \models r)$ in an attempt to say something like "if r then p-and-q implies r." The forbidden sequence of eleven symbols is a monstrosity, neither a formula of $\mathcal{L}$ nor a claim about such formulas.

It follows from Definition (5) that to show an argument $\varphi_1 \dots \varphi_k/\psi$ to be invalid, you must show that $[\varphi_1 \dots \varphi_k] \not\subseteq [\psi]$. This is achieved by producing a truth-assignment α such that $\alpha \in [\varphi_1 \dots \varphi_k]$ and $\alpha \notin [\psi]$. For example, to demonstrate that $p \not\models p \rightarrow q$, you can exhibit truth-assignment (c) of Table (3) inasmuch as $(c) \models p$ and $(c) \not\models p \rightarrow q$, hence $(c) \in [p]$ and $(c) \notin [p \rightarrow q]$. A truth-assignment like (c) is called "invalidating" for the argument $p/p \rightarrow q$. Officially:

(7) DEFINITION: A truth-assignment α is *invalidating* for an argument $\varphi_1 \dots \varphi_k$ / ψ just in case $\alpha \in [\varphi_1 \dots \varphi_k]$ and $\alpha \notin [\psi]$. Equivalently, α is invalidating just in case $\alpha \models \varphi_1, \dots, \alpha \models \varphi_k$ and $\alpha \not\models \psi$.

To illustrate the definition again, consider the argument $q \vee r$ / $q \rightarrow \neg r$. A check of Table (3) shows that (e) makes $q \vee r$ true but $q \rightarrow \neg r$ false. Hence, $(e) \in [q \vee r]$ whereas $(e) \notin [q \rightarrow \neg r]$. Thus, (e) is invalidating for $q \vee r$ / $q \rightarrow \neg r$. [So is (a), as you can verify.]

You've seen that if an argument is valid then it is impossible for the premises to be true and the conclusion false. This is the formal counterpart to the idea of "secure inference" introduced in Section 1.3. In the earlier discussion we were concerned with arguments written in English, and our explanation of secure inference relied on the vague concept of what "can be true." Equivalently, we could have framed the notion of secure inference in terms of the equally vague idea of "possibility" or "necessity."[3] Making these ideas precise for arguments expressed in English is a difficult affair. We therefore retreated to the simpler language $\mathcal{L}$, and defined validity in purely set-theoretical terms (namely, as inclusion between two sets of truth-assignments). In the logical realm there is no need to clarify terms like "possibility," even if we deployed them to build intuitions. An argument is valid just in case the meaning of its premises is included in the meaning of its conclusion, and "meanings" are themselves set-theoretical objects. Let us rejoice in such clarity! At least, let us rejoice until we begin to worry about the relation between $\mathcal{L}$ and natural language. But such worries are for another day. Right now, everything is perfectly clear.

[3]An inference from sentence A to sentence B is secure just in case it is *not possible* for A to be true and B false. Also, the inference is secure just in case it is necessarily the case that B is true if A is.

(8) EXERCISE: Test whether the argument $p \vee r \vee \neg q, q/p$ is valid.

(9) EXERCISE: Test whether $p \vee q, \neg q \models p$.

5.1.3 Validity and soundness

A valid argument need not have a true conclusion. All the inclusion $[\varphi_1 \ldots \varphi_k] \subseteq [\psi]$ buys you is the following guarantee. *If* the premises of the argument are true (that is, if Reality belongs to the truth-assignments that satisfy all of the premises) *then* the conclusion is true. Bets are off if not every premise is true. Suppose, for example, that p codes the statement "Elijah Lagat won the 2001 Boston Marathon." This is true (the race was amazingly close). Let q be "Bill Clinton finished among the top 10 in the 2001 Boston Marathon." This is false (Clinton didn't even make the top twenty). The argument $p \wedge q/q$ is valid despite its false conclusion. If $p \wedge q$ were true then so would q be. But since $p \wedge q$ is false, the truth-value of q is not constrained. In particular, it might be false (as in this example), or it could be true (as in the conclusion of the valid argument $p \wedge q/p$). The only case ruled out by the validity of an argument is that its premises be true but its conclusion false.

The guarantee offered by validity rests on the truth of *all* the premises. If even one premise of a valid argument is false then the conclusion may be false as well. For example, let p be "An Ethiopian won the men's Boston Marathon in 2001," and let q be "An Ethiopian won the women's Boston Marathon in 2001." Then the argument $p, p \leftrightarrow q/q$ is valid with false conclusion and one false premise (both winners were Kenyan so p is false whereas $p \leftrightarrow q$ is true). Had both premises been true, the falsity of the conclusion would have been impossible.

A valid argument with true premises is called *sound*. Thus, if p and q were "A Kenyan won the men's Boston Marathon in 2001," and "A Kenyan won the women's Boston Marathon in 2001," then $p, p \leftrightarrow q/q$ is not only valid but also sound. It's conclusion is true. Please pause a moment and try to say to yourself *why* a sound argument has true conclusion. (Don't read the end of this paragraph until you've given the matter some thought.) Let argument $\varphi_1 \ldots \varphi_k/\psi$

be given. If the argument is sound then the premises are true. Hence, Reality (the "true" truth-assignment) belongs to $[\varphi_1 \ldots \varphi_k]$. Since sound arguments are valid, $[\varphi_1 \ldots \varphi_k] \subseteq [\psi]$, hence Reality also belongs to $[\psi]$, which is just to say that ψ is true. That's why sound arguments have true conclusions.

(10) EXERCISE: Which of the following arguments are valid?

(a) $p \rightarrow q, \neg p / \neg q$

(b) $p \vee q, \neg p / q$

(c) $p \rightarrow q, \neg q / \neg p$

(d) $(p \wedge q) \vee r, \neg p / r$

(e) $p \leftrightarrow (q \vee r), \neg r / \neg p$

(11) EXERCISE: Let p, q, r be as follows.

p	Julius Caesar once visited Brooklyn.
q	Julius Caesar has been to Coney Island.
r	Julius Caesar ate french fries with vinegar.

Is the argument with premises $p \rightarrow (q \wedge r), \neg r$ and conclusion $\neg p$ sound?

(12) EXERCISE: Examine all relevant truth-assignments to convince yourself of the following claims.

(a) $((p \rightarrow q) \rightarrow p) \models p$

(b) $\neg(p \rightarrow q) \models (p \wedge \neg q)$

5.2 Tautology

Now that you know about validity in Sentential Logic, it's time for our next concept: *tautology*.

5.2.1 Tautologies and truth tables

Remember truth tables? (If not, please review Section 4.2.5.) Let's do a truth-table for $(q \wedge r) \rightarrow r$.

(13)

$($	q	$\wedge$	r	$)$	$\rightarrow$	r
	T	T	T		T	T
	T	F	F		T	F
	F	F	T		T	T
	F	F	F		T	F

Under the principal connective $\rightarrow$ we see a column of T's. Thus, no matter what a truth-assignment says about the truth and falsity of q and r, it satisfies $(q \wedge r) \rightarrow r$. A formula with this property is called a "tautology." Officially:

(14) DEFINITION: A formula φ is *a tautology* (or *tautologous*) just in case $[\varphi] = \mathsf{TrAs}$. If φ is a tautology, we write $\models \varphi$, otherwise $\not\models \varphi$.

Recall that TrAs is the set of all truth-assignments. So, $[\varphi] = \mathsf{TrAs}$ in Definition (14) signifies that every truth-assignment falls into the meaning of φ. Hence, Reality (the "true" truth-assignment) is *guaranteed* to fall into the meaning of φ. Hence, φ is *guaranteed* to be true. In terms of the discussion in Section 4.3.3, a formula is a tautology just in case it expresses the vacuous meaning consisting of all truth-assignments. Asserting such a formula does not circumscribe the possible realities.

Notice the new use of $\models$. If there is nothing to its left then it signifies that the formula to its right is tautologous. So the symbol $\models$ now has three missions, namely, (a) to signify satisfaction of a formula by a truth-assignment, as in $\alpha \models \varphi$, (b) to signify the validity of arguments, as in $\varphi_1 \ldots \varphi_k \models \psi$, and (c) to signify tautology, as in $\models \varphi$. Mission (c), however, can best be seen as a special case of (b). We can read $\models \varphi$ as $\emptyset \models \varphi$, thinking of $\emptyset / \varphi$ as an argument with no premises. Thus, $\models \varphi$ signifies that no premises at all are needed to guarantee the truth of φ. Tautologies are already guaranteed to be true (without the help of any premises) because every truth-assignment satisfies them. It helps to think of the matter as follows. The premises of a valid argument cut down the set of truth-assignments to a set small enough to fit into the meaning of the conclusion. When the conclusion is tautologous, there is no need to cut this set down since the conclusion embraces *all* of the truth-assignments.

Here are a some more tautologies with their truth tables. Others are left as exercises.

(15)

$(p$	$\wedge$	$q)$	$\leftrightarrow$	$(q$	$\wedge$	$p)$
T	T	T	T	T	T	T
T	F	F	T	F	F	T
F	F	T	T	T	F	F
F	F	F	T	F	F	F

(16)

$\neg$	r	$\vee$	$((r$	$\wedge$	$p)$	$\vee$	$(r$	$\wedge$	$\neg$	$p))$
F	T	T	T	T	T	T	T	F	F	T
T	F	T	F	F	T	F	F	F	F	T
F	T	T	T	F	F	T	T	T	T	F
T	F	T	F	F	F	F	F	F	T	F

(17)

$(p$	$\wedge$	$q)$	$\rightarrow$	$(p$	$\vee$	$r)$
T	T	T	T	T	T	T
T	T	T	T	T	T	F
T	F	F	T	T	T	T
T	F	F	T	T	T	F
F	F	T	T	F	T	T
F	F	T	T	F	F	F
F	F	F	T	F	T	T
F	F	F	T	F	F	F

The simplest tautologies are $p \rightarrow p$ and $p \vee \neg p$. It should take just a moment for you to verify that these formulas are indeed tautologous.

(18) EXERCISE: Which of the following formulas are tautologies? (You'll need to construct truth-tables to find out.)

(a) $(p \vee q) \rightarrow p$

(b) $p \rightarrow (p \vee q)$

(c) $p \vee (\neg p \vee q)$

(d) $(p \wedge q) \vee (p \wedge \neg q) \vee \neg p$

(e) $p \rightarrow \neg p$

(19) EXERCISE: Write out truth-tables to convince yourself that the following formulas are tautologies.

(a) $(p \rightarrow q) \vee (q \rightarrow p)$

(b) $(p \rightarrow q) \vee (q \rightarrow r)$

(c) $p \vee (p \rightarrow q)$

5.2.2 Tautologies and implication

Suppose that p represents the sentence "King Solomon was born in ancient Israel," and let q be "King Solomon did not see Episode II of *Star Wars*." Then $p \rightarrow q$ is true. (Right?) Can we pronounce this formula as "p implies q"? No. Definition (5)c reserves the word "implies" for the relation between formulas that is symbolized by $\models$. Uttering "p implies q" thus invites the interpretation $p \models q$. The latter claim is a falsehood. You can see that $p \not\models q$ by observing that $[p] \not\subseteq [q]$. The latter fact is visible from Table (3); the truth-assignment (c), for example is a member of $[p]$ but not a member of $[q]$. Someone who asserts the falsehood "p implies q" probably has in mind the truth of $p \rightarrow q$. For now, a good way to pronounce the latter formula is "if p then q." (Later we'll worry about whether *if–then–* really does justice to the arrow.[4])

Our example shows us the importance of distinguishing the claim that φ implies ψ from the claim that $\varphi \rightarrow \psi$ is true. There is nonetheless an important connection between $\rightarrow$ and $\models$. It is stated in the following fact, often called the *Deduction Theorem*.

(20) FACT: Let $\Gamma \subseteq \mathcal{L}$, and $\varphi, \psi \in \mathcal{L}$ be given.[5] Then $\Gamma \cup \{\varphi\} \models \psi$ if and only if

[4]It was in Section 3.7 that we first warned against pronouncing $\rightarrow$ as "implies."

[5]That is, let there be given a set Γ (capital "gamma") of formulas, and two specific formulas φ and ψ.

$\Gamma \models \varphi \rightarrow \psi$.[6]

To illustrate, let Γ consist of the two formulas $(p \wedge q) \rightarrow r$ and p. Let φ be q and let ψ be r. Then Fact (20) yields:

(21) $\{(p \wedge q) \rightarrow r, p, q\} \models r$ if and only if $\{(p \wedge q) \rightarrow r, p\} \models q \rightarrow r$.

If you use Table (3) to calculate $[(p \wedge q) \rightarrow r, p]$, $[(p \wedge q) \rightarrow r, p, q]$ $[q \rightarrow r]$, and $[r]$, you'll see that (21) is true because the left and the right side of the "if and only if" are both true. In fact, we get:

$[(p \wedge q) \rightarrow r, p, q]$	$\{(a)\}$
$[r]$	$\{(a), (c), (e), (g)\}$
$[(p \wedge q) \rightarrow r, p]$	$\{(a), (c), (d)\}$
$[q \rightarrow r]$	$\{(a), (c), (d), (e), (g), (h)\}$

Hence $[(p \wedge q) \rightarrow r, p, q] \subseteq [r]$ and $[(p \wedge q) \rightarrow r, p] \subseteq [q \rightarrow r]$.

For a contrasting case, let Γ be as before but switch the interpretation of φ and ψ. Now Fact (20) yields:

(22) $\{(p \wedge q) \rightarrow r, p, r\} \models q$ if and only if $\{(p \wedge q) \rightarrow r, p\} \models r \rightarrow q$.

Some more calculation of meanings reveals that both the left and right hand sides of (22) are false, so (22) itself is true. Let's see why Fact (20) is true in general.

Proof of Fact (20): There are two directions to consider. First suppose that the lefthand side of (20) is true. We must show that the right hand side of (20) is true. (Then we'll switch directions.) Our supposition is:

(23) $\Gamma \cup \{\varphi\} \models \psi$.

[6] Reminder: $\Gamma \cup \{\varphi\}$ is the set consisting of the members of Γ along with φ (as an additional member). See Section 2.5.

To prove that $\Gamma \models \varphi \to \psi$, we must show that $[\Gamma] \subseteq [\varphi \to \psi]$. So consider an arbitrary truth-assignment $\alpha \in [\Gamma]$. It suffices to show that $\alpha \in [\varphi \to \psi]$. There are two possibilities, namely, $\alpha \in [\varphi]$ and $\alpha \notin [\varphi]$. (Exactly one of these two possibilities must be the case.) Suppose first that $\alpha \in [\varphi]$. Then both $\alpha \in [\Gamma]$ and $\alpha \in [\varphi]$. Hence $\alpha \in [\Gamma \cup \{\varphi\}]$.[7] So by (23) (which implies that $[\Gamma \cup \{\varphi\}] \subseteq [\psi]$), $\alpha \in [\psi]$. By Fact (30)d of Section 4.4.2, this shows that $\alpha \in [\varphi \to \psi]$, which is what we wanted to prove. For convenience, the earlier fact is repeated here.

(24) FACT: Suppose that φ is the conditional $\chi \to \psi$. Then $[\varphi] = (\mathsf{TrAs} - [\chi]) \cup [\psi]$.

The other possibility is $\alpha \notin [\varphi]$, hence $\alpha \in \mathsf{TrAs} - [\varphi]$. Then (24) yields immediately that $\alpha \in [\varphi \to \psi]$, again yielding what we wanted to prove. So we've proved Fact (20) from left to right.

For the right-to-left direction, suppose this time that:

(25) $\Gamma \models \varphi \to \psi$.

To prove $\Gamma \cup \{\varphi\} \models \psi$, we must show that $[\Gamma \cup \{\varphi\}] \subseteq [\psi]$. So choose arbitrary $\alpha \in [\Gamma \cup \{\varphi\}]$. It must be shown that $\alpha \in [\psi]$. Since $\alpha \in [\Gamma \cup \{\varphi\}]$, $\alpha \in [\Gamma]$.[8] Hence by (25) (which implies $[\Gamma] \subseteq [\varphi \to \psi]$), $\alpha \in [\varphi \to \psi]$. By (24), we thus have that $\alpha \in (\mathsf{TrAs} - [\varphi]) \cup [\psi]$. Since $\alpha \in [\Gamma \cup \{\varphi\}]$, $\alpha \in [\varphi]$ hence $\alpha \notin \mathsf{TrAs} - [\varphi]$. Therefore, $\alpha \in [\psi]$. So we're done with the proof. (This state of affairs is marked by a black box, as follows.) ∎

Now that we've proved Fact (20), let us draw out a corollary. When Γ is the empty set, (20) becomes:

(26) FACT: Let formulas φ and ψ be given. Then $\varphi \models \psi$ if and only if $\models \varphi \to \psi$.

---

[7]To understand this step of the proof, ask yourself: What is $[\Gamma \cup \{\varphi\}]$? By Definition (2), $[\Gamma \cup \{\varphi\}]$ is the set of truth-assignments that satisfy every member of $\Gamma \cup \{\varphi\}$. This is the set of truth-assignments that satisfy all of Γ and also φ. Since both $\alpha \in [\Gamma]$ and $\alpha \in [\varphi]$, it follows that $\alpha \in [\Gamma \cup \{\varphi\}]$.

[8]By Definition (2), $[\Gamma \cup \{\varphi\}]$ is the set of truth-assignments that satisfy every member of $\Gamma \cup \{\varphi\}$, hence satisfy every member of Γ. So $\alpha \in [\Gamma \cup \{\varphi\}]$ allows us to infer that $\alpha \in [\Gamma]$.

In other words, φ implies ψ just in case $\varphi \rightarrow \psi$ is a tautology. This nice fact gives us a new means of testing whether one formula implies another. Just form their conditional and write down its truth table. For example, Table (17) showed us that $\models (p \wedge q) \rightarrow (p \vee r)$. So we may conclude from (26) that $(p \wedge q) \models (p \vee r)$. Indeed, Fact (20) has a more general corollary that relies on (4), equating $[\{\varphi_1 \ldots \varphi_k\}]$ and $[\varphi_1 \wedge \ldots \wedge \varphi_k]$. The more general version may be stated as follows.

(27) FACT: Let formulas $\varphi_1 \ldots \varphi_k$ and ψ be given. Then $\{\varphi \ldots \varphi_k\} \models \psi$ if and only if $\models (\varphi_1 \wedge \ldots \wedge \varphi_k) \rightarrow \psi$.

Thus, Table (17) also shows us that $\{p, q\} \models (p \vee r)$.

5.2.3 Implications involving tautologies

We make a few more points about implication and tautology before turning to contradiction. Let φ be your favorite tautology (we like $p \rightarrow p$). Then by (14), $[\varphi] = \mathsf{TrAs}$, the set of all truth-assignments. Let ψ be any other formula. Then $[\psi] \subseteq \mathsf{TrAs}$ (of course), so $[\psi] \subseteq [\varphi]$ (because $[\varphi] = \mathsf{TrAs}$). By Definition (5), the latter inclusion yields $\psi \models \varphi$. We conclude:

(28) FACT: For all formulas φ, ψ, if $\models \varphi$ then $\psi \models \varphi$. (Every formula implies a tautology.)

Similarly, if φ is tautological then $[\varphi] \subseteq [\psi]$ only if $[\psi] = \mathsf{TrAs}$ hence only if ψ is also a tautology. That is:

(29) FACT: For all formulas φ, ψ, if $\models \varphi$ and $\varphi \models \psi$ then $\models \psi$. (Tautologies only imply tautologies.)

Perhaps these facts seems strange to you. In that case, you'll find contradictions even stranger.

5.3 Contradiction

5.3.1 Contradictions and truth tables

Here is the truth table for $p \wedge \neg p$.

(30)

p	$\wedge$	$\neg$	p
T	F	F	T
F	F	T	F

Under the principal connective $\wedge$ we see only F, which means that no truth-assignment satisfies $p \wedge \neg p$. Such formulas express the empty meaning and are called "contradictions." Officially:

(31) DEFINITION: Let $\varphi \in \mathcal{L}$ be given. φ is *a contradiction* (or *contradictory*) just in case $[\varphi] = \emptyset$.

Here is another example, complete with truth table.

(32)

q	$\wedge$	$((r$	$\wedge$	$\neg$	$q)$	$\vee$	$(\neg$	r	$\wedge$	$\neg$	$q))$
T	F	T	F	F	T	F	F	T	F	F	T
T	F	F	F	F	T	F	T	F	F	F	T
F	F	T	T	T	F	T	F	T	F	T	F
F	F	F	F	T	F	T	T	F	T	T	F

Just as tautologies are guaranteed to be true, contradictions are guaranteed to be false. This is because Reality — the "real" truth-assignment, corresponding to the facts — can't be a member of $[\varphi]$ when φ is a contradiction (because in this case $[\varphi]$ has no members). We first brought empty meanings to your attention in Section 4.3.3.

Since the meaning of a contradiction is empty, no truth-assignment satisfies it. We therefore say that contradictions are "unsatisfiable." Officially:

(33) DEFINITION: Let formula φ be given. If $[\varphi] \neq \emptyset$ then φ is said to be *satisfiable*. Otherwise [if $[\varphi] = \emptyset$, hence φ is a contradiction], φ is said to be *unsatisfiable*.

(34) EXERCISE: Which of the following formulas are contradictions?

(a) $(p \vee q) \wedge \neg(p \wedge q)$

(b) $(p \vee \neg q) \wedge \neg p \wedge q$

(c) $p \rightarrow \neg p$

(d) $(p \wedge q) \leftrightarrow \neg(p \vee q)$

5.3.2 Contradictions and implication

To make an important point about contradiction, we need to remember the following fact from our discussion of sets (see Section 2.6).

(35) For every set B, $\emptyset \subseteq B$.

Since the meaning of a contradiction is the empty set, we have immediately from (35):

(36) FACT: Suppose that $\varphi \in \mathcal{L}$ is a contradiction. Then for every formula ψ, $\varphi \models \psi$.

For example, $p \leftrightarrow \neg p \models q \wedge r$, since $p \leftrightarrow \neg p$ is a contradiction (as you can easily check).

Contradictions imply everything. Isn't that weird? Actually, an independent proof can be given for the claim that p and not-p — the poster boy contradiction — implies any sentence ψ.[9]

(a) Suppose p and not-p are both true.

(b) From the assumption (a), p is true.

(c) From (b), at least one of p and ψ is true.

(d) From the assumption (a), not-p is true, hence p is not true.

[9] We follow the discussion in Sanford [89, p. 74].

(e) From (c) and (d), we conclude ψ. For, at least one of p and ψ is true [according to (c)], and it's not p [according to (d)]. ∎

Finally, we note an analogue to Fact (29). Just as tautologies only imply tautologies, contradictions are only implied by contradictions. We leave the proof to you.

(37) FACT: Let $\varphi, \psi \in \mathcal{L}$ be given. If ψ is a contradiction and $\varphi \models \psi$ then φ is also a contradiction.

(38) EXERCISE: Prove Fact (37).

5.3.3 Contingency

In between the tautologies and contradictions are the formulas that are satisfied by some but not all truth-assignments. Let us give them a name.

(39) DEFINITION: A formula φ is *contingent* (or, *a contingency*) just in case $\emptyset \neq [\varphi] \neq \mathsf{TrAs}$.[10]

That is, φ is contingent just in case there are truth-assignments α, β such that $\alpha \models \varphi$ and $\beta \not\models \varphi$. The truth of such a formula cannot be decided by constructing a truth table. You must consult reality and see whether the formula lies among the truth-assignments that satisfy the formula, or those that don't.

In Section 4.3.3 we characterized a meaning as contingent if it is neither $\emptyset$ nor TrAs. So, Definition (39) stipulates that a formula is contingent just in case its meaning is contingent.

There is yet another way to characterize the contingent formulas. It relies on the following fact, which is evident from the interpretation of $\neg$ [see Table (15) in Section 4.2.4].

(40) FACT: For every truth-assignment α and every formula φ, either $\alpha \models \varphi$ or $\alpha \models \neg\varphi$ (and not both).

[10] The expression $\emptyset \neq [\varphi] \neq \mathsf{TrAs}$ is shorthand for the two claims: $[\varphi] \neq \emptyset$, and $[\varphi] \neq \mathsf{TrAs}$.

It follows from (39) and (40) that:

(41) FACT: A formula φ is contingent just in case there are truth-assignments α, β such that $\alpha \models \varphi$ and $\beta \models \neg\varphi$.

From (41) it should be clear that the set of contingent formulas is *closed under negation*. By this is meant:

(42) FACT: A formula is contingent if and only if its negation is contingent.

For example, p is contingent and so are $\neg p$, $\neg\neg p$, etc. Likewise, Table (20) of Section 4.2.5 shows that $p \rightarrow (q \wedge p)$ is contingent, so $\neg(p \rightarrow (q \wedge p))$ is also contingent. What about tautologies and contradictions? Are they closed under negation? No. In fact, you can easily see that exactly the reverse is true, namely:

(43) FACT: A formula is a tautology if and only if its negation is a contradiction.

For example, the negation of the tautology $p \leftrightarrow p$ is the contradiction $\neg(p \leftrightarrow p)$. Why is (43) true in general? Well, by (30)a of Section 4.4.2, $[\neg\varphi] = \mathsf{TrAs} - [\varphi]$. So, if $[\varphi] = \mathsf{TrAs}$ then $[\neg\varphi] = \emptyset$, and if $[\varphi] = \emptyset$ then $[\neg\varphi] = \mathsf{TrAs}$.

Suppose that φ is tautologous and ψ is contingent. Can you conclude anything about their conjunction $\varphi \wedge \psi$? Could it be a contradiction? A tautology? A contingency? (You might want to consider some examples before answering.) Correct! $\varphi \wedge \psi$ must be contingent. After all, since ψ is contingent there are truth-assignments α, β such that $\alpha \models \psi$ and $\beta \not\models \psi$.[11] Since $\alpha \models \psi$ then also $\alpha \models \varphi \wedge \psi$ (since $\alpha \models \varphi$); and since $\beta \not\models \psi$, $\beta \not\models \varphi \wedge \psi$. So there are truth-assignments α, β such that $\alpha \models \varphi \wedge \psi$ and $\beta \not\models \varphi \wedge \psi$. Hence $\varphi \wedge \psi$ is contingent.

Now suppose that φ and ψ are both contingent. What about their conjunction, must it also be contingent? (Take your time; we'll stay right here.) The

[11]Reminder: We use $\beta \not\models \psi$ to mean that it is not the case that $\beta \models \psi$. See (13)a of Section 4.2.3.

simplest example solves the matter. Let φ be p and ψ be $\neg p$. Both are contingent yet their conjunction $(p \wedge \neg p)$ is a contradiction. So the correct answer to our query is No.

Aren't these fun? The following exercise offers similar problems.

(44) EXERCISE: Let formulas φ, ψ be given. Mark the following claims as true or false, and give a reason for each answer.

(a) If both φ and ψ are tautologies then so is their conjunction. (That is, if $\models \varphi$ and $\models \psi$ then $\models \varphi \wedge \psi$.)

(b) If both φ and ψ are tautologies then so is their disjunction.

(c) If both φ and ψ are contradictions then so is their conjunction.

(d) If both φ and ψ are contradictions then so is their disjunction.

(e) If both φ and ψ are contingent then so is their disjunction.

(f) If both φ and ψ are tautologies then so is $\varphi \rightarrow \psi$

(g) If both φ and ψ are contradictions then so is $\varphi \rightarrow \psi$

(h) If both φ and ψ are contingent then so is $\varphi \rightarrow \psi$

(45) EXERCISE: Test each of the following formulas for tautology, contradiction, and contingency.

(a) $p \vee \neg(p \wedge q)$

(b) $p \wedge \neg(p \vee q)$

(c) $p \rightarrow (q \wedge p)$

(d) $(p \wedge q) \rightarrow (p \vee \neg q)$

(e) $p \wedge q \rightarrow (\neg p \vee \neg q)$

(f) $(p \vee q) \wedge (\neg q \wedge \neg p)$

5.4 Logical equivalence

Logical implication is not *symmetric*. That is, it can hold in one direction without holding in the other. For example $p \wedge q \models p$ whereas $p \not\models p \wedge q$.[12] On the other hand, it may happen that the implication runs in both directions. For example, $p \vee q \models q \vee p$ and $q \vee p \models p \vee q$. In this symmetrical case, we say that the two formulas are "logically equivalent."

Now if $\varphi \models \psi$ and $\psi \models \varphi$ then Definition (5) yields $[\varphi] \subseteq [\psi]$ and $[\psi] \subseteq [\varphi]$. You know that for any two sets X, Y, $X \subseteq Y$ and $Y \subseteq X$ just in case $X = Y$.[13] So it follows that two formulas are logically equivalent just in case they have the same meaning. We use this fact to formulate our official definition of logical equivalence.

(46) DEFINITION: Formulas φ, ψ are *logically equivalent* just in case $[\varphi] = [\psi]$.

If φ and ψ are logically equivalent, we also say that φ is *logically equivalent to* ψ. For example, $p \wedge q$ is logically equivalent to $q \wedge p$ since (as easily seen) $[p \wedge q] = [q \wedge p]$. From our remarks above, we have the following fact.

(47) FACT: Formulas φ, ψ are logically equivalent if and only if $\varphi \models \psi$ and $\psi \models \varphi$.

For a revealing example of logical equivalence, let us compute $[p \leftrightarrow q]$ and $[(p \rightarrow q) \wedge (q \rightarrow p)]$. Referring to Table (3) above, and Table (19) in Section 4.2.4 (for biconditionals), we compute $[p \leftrightarrow q] = \{(a), (b), (g), (h)\}$. Turning now to $[(p \rightarrow q) \wedge (q \rightarrow p)]$, we know that a conditional is false just in case the left hand side is true and the right hand side is false; otherwise, it is true. [See Table (18) in Section 4.2.4.] A little reflection then shows that a truth-assignment satisfies $(p \rightarrow q) \wedge (q \rightarrow p)$ just in case it assigns the same truth-value to p and q. A look at (3) shows that this condition is met just for the truth-assignments $\{(a), (b), (g), (h)\}$. Therefore, $[p \leftrightarrow q] = [(p \rightarrow q) \wedge (q \rightarrow p)]$, and $p \leftrightarrow q$ is logically

[12] You can use Table (3) to verify that $[p \wedge q] \subseteq [p]$ and $[p] \not\subseteq [p \wedge q]$. (We use $\not\subseteq$ to signify that $\subseteq$ does not hold.)

[13] See Section 2.2 in case you've forgotten why this is true.

equivalent to $(p \to q) \wedge (q \to p)$. Such a nice example is worth recording in more general form:

(48) FACT: For every pair φ, ψ of formulas, $\varphi \leftrightarrow \psi$ is logically equivalent to $(\varphi \to \psi) \wedge (\psi \to \varphi)$.

Finally, we record a principle that is is analogous to Fact (26).

(49) FACT: Formulas φ, ψ are logically equivalent if and only if $\models \varphi \leftrightarrow \psi$.

Proof of Fact (49): First we go from left to right. If φ and ψ are logically equivalent then $[\varphi] = [\psi]$. So, for a given truth-assignment α, either α satisfies both φ and ψ or neither of them. Hence $\alpha \models \varphi \leftrightarrow \psi$. Since α was chosen arbitrarily, this yields $\models \varphi \leftrightarrow \psi$.

For the other direction, suppose that $\models \varphi \leftrightarrow \psi$, and let arbitrary truth-assignment α be given. Since $\alpha \models \varphi \leftrightarrow \psi$, α either satisfies both φ and ψ or neither of them. Hence, either $\alpha \in [\varphi]$ and $\alpha \in [\psi]$ or $\alpha \notin [\varphi]$ and $\alpha \notin [\psi]$. Hence (since α was chosen arbitrarily), $[\varphi] = [\psi]$. ∎

Drawing together threads of the preceding discussion, we can see that tautology, implication, contradiction, and logical equivalence are different expressions of the same concept. They can all be defined in terms of each other. The following fact summarizes the matter. You've seen most of its assertions before. Others are new. You're asked to prove the new stuff in Exercise (53).

(50) FACT: Let formulas φ and ψ be given.

(a) $\models \varphi$ if and only if for all formulas χ, $\chi \models \varphi$.

(b) φ is a contradiction if and only if for all formulas χ, $\varphi \models \chi$.

(c) φ, ψ are logically equivalent if and only if $\models \varphi \leftrightarrow \psi$.

(d) $\models \varphi$ if and only if $\neg\varphi$ is a contradiction.

(e) φ is a contradiction if and only if $\models \neg\varphi$.

(f) $\varphi \models \psi$ if and only if $\models \varphi \to \psi$.

(g) φ, ψ are logically equivalent if and only if $\models \varphi \leftrightarrow \psi$.

(h) $\varphi \models \psi$ if and only if φ is logically equivalent to $\varphi \wedge \psi$.

(i) $\varphi \models \psi$ if and only if $\varphi \wedge \neg\psi$ is a contradiction.

(51) EXERCISE: Examine all relevant truth-assignments to convince yourself of the logical equivalence of $(p \rightarrow (q \vee r))$ and $((p \rightarrow q) \vee r)$.

(52) EXERCISE: Which of the following pairs of formulas are logically equivalent?

(a) p, $\neg p \rightarrow p$

(b) $p \wedge \neg q$, $\neg(p \vee \neg q)$

(c) $p \rightarrow p$, $q \vee \neg q$

(d) $p \leftrightarrow q$, $(\neg p \vee q) \wedge (p \vee \neg q)$

(e) $p \wedge \neg p$, $q \leftrightarrow \neg q$.

(53) EXERCISE: Prove parts (h) and (i) of Fact (50).

5.5 Effability

We've covered a lot of material in this chapter, and you've been doing very well. We need you to stay focussed a little longer since the chapter ends with subtle but beautiful ideas. First we'll see that every meaning in Sentential Logic is expressed by some formula. Then we'll explain how Sentential Logic can be understood as a division ("partition") of the meanings among the formulas.

Definition (28) in Section 4.4.1 gave every formula a meaning. But what about the other direction? Does every meaning get a formula? We'll now see that the answer is affirmative.

(54) THEOREM: For every $M \in \mathsf{Meanings}$ there is $\varphi \in \mathcal{L}$ such that $[\varphi] = M$.

Recall from Definition (25) in Section 4.3.2 that Meanings is the class of all meanings. It thus consists of every subset of truth-assignments. Theorem (54) asserts that each of them is expressed by some formula. To confirm (54), let us start with a simpler fact.

(55) FACT: For every truth-assignment α there is $\varphi \in \mathcal{L}$ such that $[\varphi] = \{\alpha\}$.

To illustrate, consider (a) in Table (3) (assuming, as usual, that there are just three variables in $\mathcal{L}$). It is clear that $[p \wedge q \wedge r] = \{(a)\}$. After all, (a) satisfies $p \wedge q \wedge r$, and none of the other truth-assignments in Table (3) satisfy $p \wedge q \wedge r$ since each fails to satisfy at least one of p, q, r. For another illustration, consider (d) in (3). You can see that $[p \wedge \neg q \wedge \neg r] = \{(d)\}$. Again, it is obvious that $(d) \models p \wedge \neg q \wedge \neg r$, and equally obvious that no other truth-assignment satisfies this formula. For, every other truth-assignment fails to satisfy at least one of p, $\neg q$, $\neg r$. This should be enough to convince you of Fact (55).

What about a pair of truth-assignments, α, β? Is there a formula that expresses $\{\alpha, \beta\}$? Sure. Consider $\{(a), (d)\}$. It is "meant" by $(p \wedge q \wedge r) \vee (p \wedge \neg q \wedge \neg r)$. On the one hand, it is clear that $\{(a), (d)\} \subseteq [(p \wedge q \wedge r) \vee (p \wedge \neg q \wedge \neg r)]$ since each of (a) and (d) satisfy exactly one disjunct of this formula. On the other hand, $[(p \wedge q \wedge r) \vee (p \wedge \neg q \wedge \neg r)]$ includes nothing more than $\{(a), (d)\}$. For, every other truth-assignment satisfies neither disjunct of $(p \wedge q \wedge r) \vee (p \wedge \neg q \wedge \neg r)$.

More generally, suppose that $\mathcal{L}$ contains n variables, $v_1, v_2 \ldots v_n$, and let truth-assignment α be given. We use $\varphi(\alpha)$ to denote the conjunction $\pm v_1 \wedge \pm v_2 \wedge \ldots \wedge \pm v_m$, where the $\pm$ sign next to v_i is replaced by a blank if $\alpha \models v_i$, and is replaced by $\neg$ if $\alpha \models \neg v_i$.[14] For example, appealing again to Table (3), we have that $\varphi(b) = p \wedge q \wedge \neg r$. Since, $[p \wedge q \wedge \neg r] = \{(b)\}$, it thus follows that $[\varphi(b)] = \{(b)\}$. Indeed, you can see that $[\varphi(\alpha)] = \{\alpha\}$ for any truth-assignment α. Similarly, it is now clear that given any set $\{\alpha_1 \ldots \alpha_m\}$ of truth-assignments, we have that:

$$[\varphi(\alpha_1) \vee \varphi(\alpha_2) \ldots \vee \varphi(\alpha_m)] = \{\alpha_1 \ldots \alpha_m\}.$$

[14] Take a peek at Fact (40) to recall that $\alpha \not\models v_i$ if and only if $\alpha \models \neg v_i$.

To illustrate, consider the set $\{(a),(d),(g)\}$ from Table (3). We've already seen what $\varphi(a)$ and $\varphi(d)$ are. You can verify that $\varphi(g)$ is $\neg p \wedge \neg q \wedge r$. So:

$$[(p \wedge q \wedge r) \vee (p \wedge \neg q \wedge \neg r) \vee (\neg p \wedge \neg q \wedge r)] = \{(a),(d),(g)\}.$$

We take it that you are now convinced that every set of truth-assignments is meant by some formula. Since meanings are nothing but sets of truth-assignments, we have thus proved Fact (54).

(56) EXERCISE: Write a formula φ such that for all truth-assignments α, $\alpha \models \varphi$ if and only if either $\alpha \models p$ or $\alpha \models q$ *but not both*. Such a formula φ expresses the *exclusive disjunction* of p and q (in contrast to the *inclusive disjunction* $p \vee q$).

5.6 Disjunctive normal form

So, every meaning is expressible in our language. But there's something even better. For every meaning there is a *nice* formula that expresses it, like the one appearing in the last example (expressing $\{(a),(d),(g)\}$). We call such formulas "nice" because they have a nice property that will be exploited in Chapter 7. In the present section we'll indicate precisely the kind of formula we have in mind. This will take several definitions. As a preliminary, we remind you about long conjunctions and disjunctions, discussed in Section 4.4.3. Because $[p \wedge (q \wedge r)] = [(p \wedge q) \wedge r)]$, we agreed earlier that the parentheses could be dropped. So let us further agree to use the term "conjunction" to denote any formula of the form $\varphi_1 \wedge \cdots \wedge \varphi_n$, where the φ_i's are arbitrary formulas. Likewise, "disjunctions" will denote any formula of the form $\varphi_1 \vee \cdots \vee \varphi_n$, for arbitrary φ_i.

(57) DEFINITION: By a *simple conjunction* is meant a variable by itself, a negated variable by itself or a conjunction (in our broader sense) of variables and negated variables.

For example, q, $\neg p$, $r \wedge \neg r \wedge q \wedge p \wedge q$ are all simple conjunctions. In contrast, neither $(r \rightarrow p) \wedge q)$ nor $\neg\neg p$ is a simple conjunction. To make sense of this definition, you'll find it helpful to think of q and $\neg p$ as conjunctions with just one conjunct. Then a simple conjunction is a conjunction whose conjuncts are variables or negated variables.

(58) DEFINITION: A formula is in *disjunctive normal form* just in case it is either a simple conjunction or a disjunction (again, in our broader sense) of simple conjunctions. We abbreviate the expression "disjunctive normal form" to *DNF*.

To illustrate the definition, consider the following formulas.

(59) EXAMPLE:

- (a) r
- (b) $r \wedge \neg p \wedge \neg q \wedge t$
- (c) $(p \wedge \neg q) \vee (r \wedge t)$
- (d) $(q \wedge \neg q) \vee (r \wedge t)$
- (e) $(p \wedge \neg q) \vee (r \wedge t \wedge q)$
- (f) $(p \wedge \neg p) \vee (r \wedge t \wedge \neg r)$
- (g) $p \vee q \vee r \vee \neg s \vee (p \wedge t)$
- (h) $(\neg\neg p \wedge \neg q) \vee (r \wedge t)$
- (i) $(p \wedge \neg q) \vee ((r \rightarrow s) \wedge t)$

Formulas (59)a - g are all in DNF. Formula (59)h is not in DNF because $\neg\neg p$ is not a simple conjunction. Similarly, (59)i is not in DNF because $r \rightarrow s$ is not a simple conjunction. To get a grip on cases (59)a,b, think of them as disjunctions with just one disjunct.

Our proof of Theorem (54) in Section 5.5 showed that for every meaning we can construct a formula in DNF with that meaning. You can verify this claim by contemplating the form of $\varphi(\alpha_1) \vee \varphi(\alpha_2) \ldots \vee \varphi(\alpha_m)$, which we used to "mean" the set $\{\alpha_1 \ldots \alpha_m\}$. It is necessary, however, to consider the special case of the empty meaning, $\emptyset$. To express $\emptyset$, we cannot use formulas of form

$\varphi(\alpha)$. Fortunately, $p \wedge \neg p$ is in disjunctive normal form since it is a simple conjunction. And of course, $[p \wedge \neg p] = \emptyset$. So the proof of Theorem (54), along with consideration of $\emptyset$, yield the following corollary.

(60) COROLLARY: For every $M \in \mathsf{Meanings}$ there is a formula ψ in disjunctive normal form such that $[\psi] = M$. Hence, for every $M \in \mathsf{Meanings}$ there is a formula ψ in which just $\neg$, $\wedge$, and $\vee$ occur such that $[\psi] = M$.

The second part of Corollary (60) tells us that the connectives $\neg$, $\wedge$, $\vee$ by themselves suffice to express all meanings in Sentential Logic. Indeed, just $\neg$ and one of $\wedge$, $\vee$ is enough for this purpose since $\varphi \wedge \psi$ is logically equivalent to $\neg(\neg\varphi \vee \neg\psi)$, and $\varphi \vee \psi$ is logically equivalent to $\neg(\neg\varphi \wedge \neg\psi)$. So, for example, instead of using $p \vee (q \wedge \neg r)$ to express $[p \vee (q \wedge \neg r)]$ we can use the logically equivalent $\neg(\neg p \wedge \neg(q \wedge \neg r)$, which does not involve $\vee$. Alternatively, we could the logically equivalent $p \vee \neg(\neg q \vee r)$, which does not involve $\wedge$.

Returning to DNF, Corollary (60) immediately yields:

(61) COROLLARY: Every formula is logically equivalent to a formula in disjunctive normal form (hence, to a formula whose connectives are limited to $\neg$, $\wedge$, and $\vee$).

You are now ready to appreciate an essential fact about DNF formulas. Its formulation relies on the following definition.

(62) DEFINITION: By a *contradictory simple conjunction* is meant any conjunction that includes conjuncts of the form v and $\neg v$ for some variable v.

For example, $q \wedge r \wedge \neg t \wedge \neg r$ is a contradictory simple conjunction. The conjunction $q \wedge r \wedge \neg t \wedge \neg s$ is not a contradictory simple conjunction. It's easy to see that the meaning of a contradictory simple conjunction is empty. That is, no truth-assignment satisfies a contradictory simple conjunction. On the other hand, any simple conjunction that is not a contradictory simple conjunction is satisfiable.[15] To illustrate, $q \wedge r \wedge \neg t \wedge \neg s$ is satisfied by any truth-assignment that assigns T to q and r, and assigns F to t and s. Since a formula φ in DNF is a disjunction of simple conjunctions, if at least one disjunct is not a contradictory simple conjunction then φ is satisfiable. (This is because a disjunction is satisfied by any truth-assignment that satisfies at least one of its disjuncts.) And a formula in DNF is unsatisfiable if all of its disjuncts are contradictory simple conjunctions. The following examples will help you see why this is true. Consider the DNF formula

$$(\neg p \wedge p \wedge r) \vee (t \wedge \neg p \wedge q) \vee (\neg q \wedge p) \vee (\neg q \wedge q) \vee q.$$

Its disjuncts are the simple conjunctions $(\neg p \wedge p \wedge r)$, $(t \wedge \neg p \wedge q)$, $(\neg q \wedge p)$, $(\neg q \wedge q)$, and q. Not all of these simple conjunctions are contradictory simple conjunctions (specifically, the second, third, and fifth disjuncts are not contradictory simple conjunctions). And you can see that the formula is satisfiable. For example, any truth-assignment that makes q true satisfies the formula (because it makes the last disjunct true). Similarly, any truth-assignment that makes q false and p true satisfies the formula (because it makes the third disjunct true). Compare the DNF formula:

$$(\neg p \wedge p \wedge r) \vee (\neg t \wedge \neg q \wedge q) \vee (\neg r \wedge r) \vee (\neg q \wedge q).$$

This new formula is unsatisfiable. For a disjunction to be satisfied by a truth-assignment, at least one of its disjuncts must be made true. But all of the disjuncts in the foregoing formula are contradictory simple conjunctions. These examples should suffice to convince you that:

(63) FACT: Let φ be a formula $\chi_1 \vee \chi_2 \vee \cdots \vee \chi_n$ in DNF. Then φ is unsatisfiable if and only if for all $i \leq n$, χ_i is a contradictory simple conjunction.[16]

(64) EXERCISE: Suppose (as usual) that our variables are limited to p, q, r. Write formulas in disjunctive normal form that express the following meanings.

---

[15] For "satisfiable," see Definition (33) in Section 5.3.1.

[16] The expression "for all $i \leq n$," means "for each of $1, 2 \ldots n$."

(a) $\{(a), (b), (c), (h)\}$.

(b) $\{(b), (c), (h)\}$.

(c) $\{(b), (d), (b)\}$.

(d) $\{(a), (b), (c), (d), (e), (f), (g), (h)\}$.

5.7 Partitioning $\mathcal{L}$ on the basis of meaning

We have seen that every meaning is expressed by some formula. There is nothing *ineffable* about Sentential Logic.[17] Indeed, each meaning gets "effed" by an infinity of formulas. This is a consequence of (54) and the following fact.

(65) FACT: For every formula φ there is an infinite collection of formulas ψ with $[\varphi] = [\psi]$. In other words, for every formula there are infinitely many formulas logically equivalent to it.

A cheap way to construct an infinite collection of formulas that mean what φ does is to add $2 \times m$ occurrences of $\neg$ in front of φ for all $m > 1$. Another cheap trick is to conjoin φ with itself m times. There are also many other formulas that are not so transparently equivalent to φ, such as $(p \rightarrow p) \rightarrow \varphi$. So you see that (65) is true.

Do you remember how many meanings there are? We discussed this in Section 4.3.2 [see Fact (26)]. For n variables there are 2^{2^n} members of $\mathsf{Meanings}$. This number is often large, but it is finite. For each of these meanings, M, let $\mathcal{L}(M)$ be the collection of formulas that mean it. Officially:

(66) DEFINITION: For every set M of truth-assignments, $\mathcal{L}(M)$ denotes the set of formulas φ with $[\varphi] = M$.

Are you losing your grip on all this notation? Let's review. Given $\varphi \in \mathcal{L}$, $[\varphi]$ is a collection of truth-assignments. That is, the operator $[\cdot]$ maps formulas

[17] According to the dictionary [1], something is ineffable if it is "beyond expression." In contrast to Sentential Logic, many meanings remain ineffable in stronger logics, like the *predicate calculus*. See, for example, [26, §VI.3].

to their meanings. Given $M \subset \mathsf{TrAs}$, $\mathcal{L}(M)$ is a collection of formulas. That is, the operator $\mathcal{L}(\cdot)$ maps meanings to the formulas that express them. We can illustrate with two extreme cases. $[p \wedge \neg p]$ is the empty set of truth-assignments and $[p \vee \neg p]$ is the entire set of truth-assignments. Thus, $[p \wedge \neg p]$ and $[p \vee \neg p]$ are both finite sets; the first has no elements, the second has 2^n elements if there are n variables. In contrast, both $\mathcal{L}(\emptyset)$ and $\mathcal{L}(\mathsf{TrAs})$ are infinite sets; the first contains all the formulas that are contradictions ($p \wedge \neg p$, $p \wedge \neg\neg\neg p$, etc.), the second contains all the formulas that are tautologies ($q \leftrightarrow q$, $\neg q \leftrightarrow \neg q$, etc.).

We claim:

(67) FACT: The sets $\{\mathcal{L}(M) \mid M \in \mathsf{Meanings}\}$ constitute a partition of $\mathcal{L}$. There are only finitely many equivalence classes in this partition. A given equivalence class has infinitely many members. Each member of a given equivalence class is logically equivalent to the other members.[18]

The fact follows easily from what you already know. By (54), each set $\mathcal{L}(M)$ is nonempty. It is infinite by (65). No two sets of form $\mathcal{L}(M_1)$ and $\mathcal{L}(M_2)$ can intersect since each formula has just one meaning. And every formula is in some set of form $\mathcal{L}(M)$ since every formula means something. [The last two assertions follow from Definition (28) in Section 4.4.1.]

Fact (67) can be pictured as follows. The finite set of meanings is sprinkled onto the floor. Then each formula is placed on the meaning it expresses. When the job is done, every formula will be placed on exactly one meaning, and every meaning will be covered by an infinite pile of formulas each logically equivalent to the others. Each of the meanings, of course, is a set of truth-assignments, and there will be relations of inclusion among various of these sets. By Definition (5), the inclusions represent logical implication between the formulas in the associated equivalence classes.

This picture is so fundamental that it is worthwhile describing it again. Formulas are assembled into a finite number of sets (equivalence classes) each with infinitely many members all logically equivalent to each other. For every

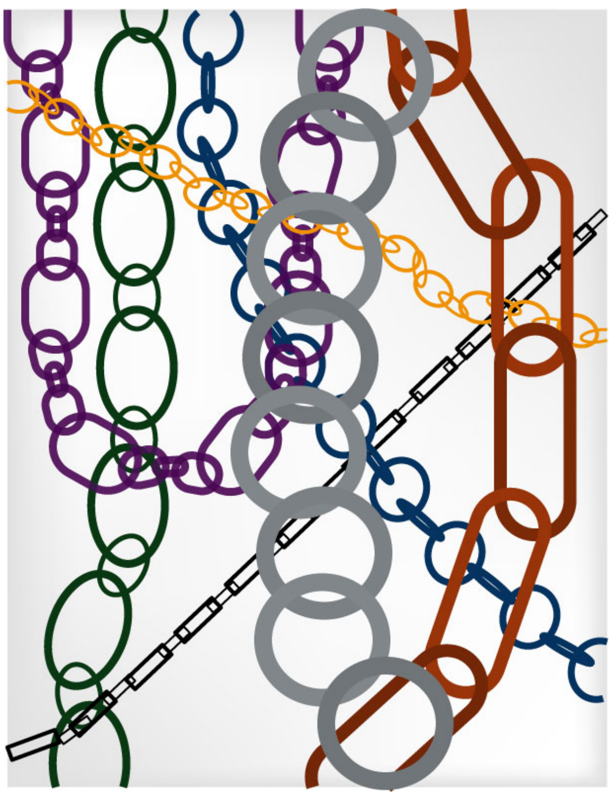
[18]For the idea of a partition of a set, see Section 2.8.

pair φ, ψ of formulas, $\varphi \models \psi$ just in case the meaning that defines φ's equivalence class is a subset of the meaning that defines ψ's equivalence class. Sentential Logic is seen thereby through the lens of meaning. Meanings imply one another (via subset relations) independently of the language $\mathcal{L}$. The latter serves only to express meanings, each getting infinitely many formulas for this purpose. It is Fact (67) that puts everything in its proper place.

This is a great picture of logic, and it culminates our study of the semantics of $\mathcal{L}$. But the picture doesn't show everything. The idea of "derivations" between formulas remains to be filled in. Derivations leave semantics to one side, relying on just the geometry of formulas to provide insight into inference. But then at the very end, semantics makes a dramatic return to the scene. It's a riveting story, and we won't spoil it now by telling you how things turn out. Instead, we invite you to join us for the next act. *Dames et Messieurs!* Seating is now available for Chapter 6.

Chapter 6

Derivation within Sentential Logic



6.1 The derivation project

What do you think? Is the following argument valid?

(1)

Premises:	$(q \vee p) \rightarrow (r \wedge s), \quad t \wedge u \wedge q$
Conclusion:	$s \vee v$

"No problem," (we hear you saying). "Relying on Definition (5) of the unforgettable Chapter 5, I'll just determine whether $[(q \vee p) \rightarrow (r \wedge s), t \wedge q]$ is a subset of $[s \vee u]$." But now you notice that the argument involves 7 variables, hence $2^7 = 128$ truth-assignments. Your enthusiasm for verifying $[(q \vee p) \rightarrow (r \wedge s), t \wedge q] \subseteq [s \vee u]$ begins to wane. "Sifting through a zillion truth-assignments is a pain. Surely there is some other way of checking validity," you muse.

Yes, there are many other ways. Various schemes have been devised for determining an argument's validity by examining just a fraction of the truth-assignments composing the meaning of its premises and conclusion. See [54] to get started, and [90] for a more comprehensive account. For our part, we are going to tell you about an entirely different approach to cumbersome arguments like (1). It is based on the idea of a *chain of reasoning*. The chain leads by small steps from the premises to the conclusion. Each step must be justified by a *rule of derivation*, to be described. It will turn out that there is a chain of justified steps from premises to conclusion just in case the argument is valid. Some of the rules of derivation resemble these:

(2) (a) Write any premise of the argument wherever you want in the chain.

(b) If a conjunction $\varphi \wedge \psi$ occurs in the chain, then you can extend the chain to the conjunct ψ.

(c) If a formula φ occurs in the chain then you can extend the chain to the disjunction $\varphi \vee \psi$.

(d) If both the formulas φ and $\varphi \rightarrow \psi$ occur in the chain then you can extend the chain to ψ.

(e) If both the formulas φ and ψ occur in the chain then you can extend the chain to the conjunction $\varphi \wedge \psi$.

Such rules have the important property of *preserving the set of truth-assignments that make the premises true*. That is, if truth-assignment α satisfies the premises of an argument then α also satisfies every member of the chain built using the premises and the rules (2). It follows that you can't get to a false conclusion by applying (2) to true premises. We'll consider truth preservation more rigorously in the next chapter. For now, it is enough to agree that the rules (2) only allow "safe" links to be added. Consider (2)b, for example. If our chain already includes a conjunction, there seems to be no risk in extending it to include the right-hand conjunct of the conjunction. For, if all the formulas in the chain were true prior to applying (2)b, they will still all be true afterwards. This is because $\varphi \wedge \psi \models \psi$, which ensures the truth of ψ assuming the truth of $\varphi \wedge \psi$. Similar remarks apply to the other rules in (2).

We can use the rules to give a rough idea of what a chain for (1) might look like, namely:

(3)

a)	$t \wedge u \wedge q$	a premise of the argument
b)	$u \wedge q$	rule (2)b applied to (a)
c)	q	rule (2)b applied to (b)
d)	$q \vee p$	rule (2)c applied to (c)
e)	$(q \vee p) \rightarrow (r \wedge s)$	another premise of the argument
f)	$r \wedge s$	rule (2)d applied to (d) and (e)
g)	s	rule (2)b applied to (f)
h)	$s \vee v$	rule (2)c applied to (g)

The chain persuades us of the validity of (1) because (2)a was exploited only to introduce premises of (1) (nothing extraneous was added). Since the remaining rules are truth preserving, all the formulas in (3) are true provided the premises of (1) are true. So, in particular, the last line of (3) is true provided the premises of (1) are true. But the last line of (3) is the conclusion of (1). Hence, the conclusion of (1) is true provided that the premises of (1) are true. The foregoing reasoning requires no assumptions about which truth-assignments make the premises of (1) true; whichever they are, we see that they also make the conclusion true. Hence, the chain (3) establishes that every truth-assignment satisfying the premises of (1) also satisfies the conclusion. In other words, (1) is valid.

The attractive feature of (3) is its brevity, compared to flipping through 128 truth-assignments. In fact, the derivation rules that we'll present allow us to sidestep truth-assignments whenever the argument under scrutiny is valid. Unfortunately, the technique won't be so handy when confronted with an invalid argument. There *is* a way of exploiting a chain to locate a truth-assignment that satisfies the premises of an invalid argument and falsifies the conclusion, thereby finding an invalidating truth-assignment.[1] This method of demonstrating invalidity is a little tedious, however, so that finding an invalidating truth-assignment is often achieved more quickly some other way than via derivation rules (such as sifting through all of the truth-assignments, looking for one that satisfies the premises but not the conclusion). So now you will surely ask:

(4) "Given an argument A, how do I know whether to (i) try to construct a chain that shows A's validity or (ii) try to find an invalidating truth-assignment? Don't I have to know in advance whether A is valid before embarking on a demonstration of its logical status?"

Well, no. You don't have to know A's logical status ahead of time. If you are willing to examine all the relevant truth-assignments, then you can announce A to be invalid if you reach a truth-assignment that satisfies the premises but not the conclusion, and you can announce A to be valid if you get to the end of the truth-assignments without finding any such example. If you want to enjoy the efficiency offered by derivations, however, then you'll need to attack A on two fronts simultaneously. You'll have to devote time to *both* the enterprises (i) and (ii) mentioned in query (4). If your hunch is that A is valid then you'll spend more time on (i) than (ii); otherwise, the reverse. And, of course, your hunch might be wrong.

There is worse to come. Even if we tell you that A is valid, it may not be evident how to build a succinct chain of steps that demonstrates A's validity. We will ultimately show that A's validity guarantees that we can find *some* chain; but we won't present a method for finding a relatively *short* chain. The

[1] For the concept of an invalidating truth-assignment, see Definition (7) in Section 5.1.2.

only methods we know for finding short chains are unbearably clumsy.[2] You'll often have to rely on ingenuity and insight to find succinct rule-based demonstrations of validity.

So Logic is not a cookbook with recipes for every dish. Sometimes a good meal depends on the creativity of the chef (that's you). But the mental effort is often worth it. A well-wrought chain of reasoning is an object of beauty. Are you ready for the esthetic challenge? Our first task is to replace talk of "chains" with a subtler idea. Chains like (3) don't suffice for our purposes because reasoning often proceeds by means of *temporary assumptions*. The next section introduces this idea informally. Then we get down to business.

Courage! This is the hardest chapter of the book.

6.2 Assumptions

Consider these two arguments, both valid.

(5) (a)

Premises	Conclusion
$p, \quad p \rightarrow (q \rightarrow r)$	$q \rightarrow r$

(b)

Premises	Conclusion
$p, \quad q \rightarrow (p \rightarrow r)$	$q \rightarrow r$

Similar, aren't they? The first can be handled by a brief chain:

a)	p	a premise of the argument
b)	$p \rightarrow (q \rightarrow r)$	another premise of the argument
c)	$q \rightarrow r$	rule (2)d applied to (a) and (b)

But our rules don't allow a chain to be built for (5)b. (Try.) Informally, the natural way to reason about the validity of (5)b is something like this:

[2]Our system is no worse than others in this regard. It can be shown that if a widely held mathematical conjecture is indeed true then there is *no* quick method for finding simple proofs of valid arguments in Sentential Logic. See [105].

> "We are given p and $q \rightarrow (p \rightarrow r)$ as premises. Let's *assume* q temporarily. That gives us $p \rightarrow r$ by applying rule (2)d to q and $q \rightarrow (p \rightarrow r)$. We then get r by applying rule (2)d to p and $p \rightarrow r$. From the assumption q (plus the premises), we have thus reached r. So $q \rightarrow r$ follows from the premises."

The rules introduced below for handling conditionals will formalize this kind of reasoning. The derivation of (5)b will go something like this:

a)	p	a premise of the argument
b)	$q \rightarrow (p \rightarrow r)$	another premise of the argument
c)	q	an *assumption* that we make temporarily.
d)	$p \rightarrow r$	rule (2)d applied to (b) and (c)
e)	r	rule (2)d applied to (a) and (d)
f)	$q \rightarrow r$	a new rule to be introduced later, applied to (c) - (e)

For other arguments we will need more than one assumption. For example, to demonstrate the validity of the argument $((p \wedge q) \rightarrow r)\ /\ p \rightarrow (q \rightarrow r)$, we will write something like this:

a)	$(p \wedge q) \rightarrow r$	premise
b)	p	temporary assumption
c)	q	another temporary assumption
d)	$p \wedge q$	(2)e applied to (b) and (c)
e)	r	rule (2)d applied to (a) and (d)
f)	$q \rightarrow r$	the new rule, applied to (c) - (e)
g)	$p \rightarrow (q \rightarrow r)$	the new rule, applied to (b) - (f)

This all looks easy but we have to be careful about assumptions lest we fall into fallacy. For example, the following chain makes careless use of an assumption to "establish" the invalid argument with premise p and conclusion $p \rightarrow (p \wedge q)$.[3]

[3] Before proceeding, you might wish to verify the invalidity of $p\ /\ p \rightarrow (p \wedge q)$ by finding an invalidating truth-assignment for it.

a)	p	premise
b)	q	temporary assumption
c)	$(p \wedge q)$	(2)e applied to (a) and (b)
d)	$p \rightarrow (p \wedge q)$	the new rule, misapplied to (a) - (c)

To avoid such mischief, our rules for conditionals will need to keep track of assumptions. For this purpose, we'll mark each assumption with the symbol $\circ$, changing it to $\bullet$ when the assumption has played its destined role.

6.3 Writing derivations

So much for informal motivation. Let's get into it.

6.3.1 The form of a successful derivation

By a *line* is meant a formula followed by one of three *marks*. The three marks are $\bullet$, $\circ$, and the blank symbol . Some lines are as follows.

$p \rightarrow q$	$\bullet$
$r \wedge t$	$\circ$
$q \vee \neg p$	
$r \leftrightarrow t$	$\bullet$

You should pronounce the mark $\circ$ as "assumption," and the mark $\bullet$ as "cancelled assumption." The intuitive significance of these labels will emerge as we proceed. The blank mark needs no name.

We now define the idea of "a derivation of an argument Γ / γ." The premises of the argument in question are represented by Γ; the conclusion is represented by γ.[4] We proceed by first defining the more general concept of a "derivation." A derivation of Γ / γ is then taken to be a derivation that meets certain conditions. Here's the definition of derivation.

[4]The symbols γ and Γ are pronounced as "gamma" and "capital gamma," respectively.

(6) DEFINITION:

(a) A *derivation* is a column of lines that is created by application of the rules explained in the next six subsections. (Rules can be used any number of times, and in any order.)

(b) A *derivation of the argument* Γ / γ is a derivation with the following properties.

i. The column ends with the line “γ $\circ$” or “γ ”.

ii. The only lines in the column which are marked by $\circ$ have formulas that are members of Γ.

If the conditions in (6)b are met, we say that the γ is *derivable* from Γ, that the argument Γ / γ is *derivable*, and similarly for other locutions. Another way to state condition (6)bi is that the column ends with a line whose formula is γ and is not marked by $\bullet$.

To explain the rules invoked in (6)a, suppose that you've already completed part of your derivation of γ from Γ. Let $\boldsymbol{D}$ be the part you've already completed. If you're just getting the derivation underway then the part you've completed is empty. In this case, $\boldsymbol{D} = \emptyset$. Otherwise, $\boldsymbol{D}$ consists of lines that were already created by application of the rules about to be introduced. Since you'll get totally lost if you don't remember what $\boldsymbol{D}$ represents, let's frame the matter.

> We use $\boldsymbol{D}$ to symbolize the part of the derivation of γ from Γ that has already been completed. (It follows that if nothing has yet been completed then $\boldsymbol{D} = \emptyset$.)

6.3.2 Assumption rule

Here is the first rule.

(7) ASSUMPTION RULE: For any formula φ whatsoever you may append the line “φ $\circ$” to the end of $\boldsymbol{D}$.

To illustrate the rule, suppose that $\boldsymbol{D}$ is

$(p \vee q)$	
p	$\circ$
r	

Then rule (7) allows you to extend $\boldsymbol{D}$ to any of the following derivations (among others).

$(p \vee q)$	$(p \vee q)$	$(p \vee q)$	$(p \vee q)$	$(p \vee q)$
p $\circ$	p $\circ$	p $\circ$	p $\circ$	p $\circ$
r	r	r	r	r
s $\circ$	$(s \leftrightarrow t)$ $\circ$	$(p \vee q)$ $\circ$	$\neg p$ $\circ$	$(\neg r \wedge (u \leftrightarrow \neg v))$ $\circ$

Notice the presence of $\circ$ in the last lines. Rule (7) can't be used without marking the line accordingly.

It's time for your first derivation! Suppose that $\Gamma = \{q, r\}$ and $\gamma = r$. If we let $\boldsymbol{D} = \emptyset$ then

$$r \quad \circ$$

extends $\boldsymbol{D}$ by an application of (7). This one-line column is a genuine derivation of r from $\{q, r\}$. You can see that it meets the conditions of Definition (6)b since (a) it ends with a line whose formula is r without the mark $\bullet$, and (b) the only lines in the derivation marked by $\circ$ comes from the set of premises. The foregoing derivation of r from $\{q, r\}$ is *also* a derivation of r from $\{p, q, r\}$. Don't let this ambiguity disturb you. A derivation of an argument $\Gamma \;/\; \gamma$ will also be a derivation of any argument $\Sigma \;/\; \gamma$ where $\Sigma \supseteq \Gamma$.[5]

6.3.3 Conditional rules

Here are two more rules for extending a derivation $\boldsymbol{D}$, both involving conditionals.

[5]Recall that $A \supseteq B$ means $B \subseteq A$; see Section 2.2. By the way, Σ is pronounced "capital sigma."

(8) CONDITIONAL INTRODUCTION RULE: Suppose that $\boldsymbol{D}$ ends with a line whose formula is ψ. Let "θ $\circ$" be the last line in $\boldsymbol{D}$ marked with $\circ$. (If there is no line in $\boldsymbol{D}$ marked with $\circ$ then you cannot use this rule.) Then you may do the following. ***First***, from "θ $\circ$" to the end of $\boldsymbol{D}$, change the mark of every line to $\bullet$ (if the mark is not $\bullet$ already). ***Next***, append the line "$\theta \rightarrow \psi$ " to the end of $\boldsymbol{D}$.

(9) CONDITIONAL ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula θ and a line with the formula $(\theta \rightarrow \psi)$ (in either order). Suppose also that neither of these lines bears the mark $\bullet$. Then you may append the line "ψ " to the end of $\boldsymbol{D}$.

To discuss these rules (and others to follow) it will be convenient to annotate our derivations as we build them. We'll number lines at the left, and explain their provenance by comments at the right. Let's use this apparatus to build a derivation, step by step, for the argument $p \rightarrow (q \rightarrow r)$ / $q \rightarrow (p \rightarrow r)$. We start with $\boldsymbol{D} = \emptyset$. Then, relying on Rule (7) (the Assumption Rule) we write:

(10)

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)

Now letting $\boldsymbol{D}$ be the derivation (10), we extend it to

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	q	$\circ$	Assumption (7)

then to:

(11)

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	q	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)

Everything should be clear up to this point. It is also clear that (11) can be extended to

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	q	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)
4	$q \rightarrow r$		Conditional Elimination Rule (9) applied to 1 and 3

and thence to:

(12)

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	q	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)
4	$q \rightarrow r$		Conditional Elimination Rule (9) applied to 1 and 3
5	r		Conditional Elimination Rule (9) applied to 2 and 4

Allow us, please, to abbreviate "Conditional Elimination Rule" to "→ elimination," and similarly for other rules to follow. We can then rewrite (12) as:

(13)

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	q	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)
4	$q \rightarrow r$		$\rightarrow$ elimination (9) on 1 and 3
5	r		$\rightarrow$ elimination (9) on 2 and 4

Now we get to see the → introduction rule (8) in action. We apply it to (13) to reach:

(14)

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	q	$\circ$	assumption (7)
3	p	$\bullet$	assumption (7)
4	$q \rightarrow r$	$\bullet$	$\rightarrow$ elimination (9) on 1 and 3
5	r	$\bullet$	$\rightarrow$ elimination (9) on 2 and 4
6	$p \rightarrow r$		$\rightarrow$ introduction (8) on 3 - 5

You should examine the transition from (13) to (14) with care. The last line in (13) marked with $\circ$ is 3. This is why all the lines from 3 to 5 are marked with $\bullet$ in (14). Notice also that the conditional introduced at line 6 has the formula at 3 as left hand side and the formula at 5 as right hand side. This is dictated by the use of rule (8) on 3 - 5. One more use of the rule suffices to finish our derivation:

(15)

1	$p \to (q \to r)$	$\circ$	Assumption (7)
2	q	$\bullet$	assumption (7)
3	p	$\bullet$	assumption (7)
4	$q \to r$	$\bullet$	$\to$ elimination (9) on 1 and 3
5	r	$\bullet$	$\to$ elimination (9) on 2 and 4
6	$p \to r$	$\bullet$	$\to$ introduction (8) on 3 - 5
7	$q \to (p \to r)$		$\to$ introduction (8) on 2 - 6

The marks on lines 3 -5 have not changed from (14) to (15) because there is no need to add a second $\bullet$. On the other hand, the $\circ$ in line 2 has been switched to $\bullet$ in (15), and line 6 has also gained a $\bullet$. But 7 is left blank, as dictated by rule (8). The three clauses of Definition (6) are now satisfied. The derivation (15) ends with unmarked $q \to (p \to r)$, the mark $\circ$ only appears next to the premise $p \to (q \to r)$, and (15) was built according to our rules. The argument $p \to (q \to r)$ / $q \to (p \to r)$ has thus been shown to be derivable.

Let's try something trickier. We'll find a derivation for $\emptyset$ / $(p \to q) \to ((q \to r) \to (p \to r))$. This is an argument with no premises.[6] How do we start to find a derivation for this argument? It is often helpful to think about what the last step of the derivation might be. If we can get $((q \to r) \to (p \to r))$ from the assumption $p \to q$ then the derivation would be completed by an application of $\to$ introduction. So let's start with:

1	$p \to q$	$\circ$	Assumption (7)

Our problem is thus reduced to deriving $((q \to r) \to (p \to r))$, and since this is a conditional we can repeat our strategy of assuming the left hand side:

1	$p \to q$	$\circ$	Assumption (7)
2	$q \to r$	$\circ$	Assumption (7)

Now we are after $p \to r$, so falling back on our strategy one more time, we arrive at:

[6] The possibility of zero premises was observed in Section 5.1.1.

1	$p \to q$	○	Assumption (7)
2	$q \to r$	○	Assumption (7)
3	p	○	assumption (7)

We're now in a position to apply $\to$ elimination to obtain:

1	$p \to q$	○	Assumption (7)
2	$q \to r$	○	Assumption (7)
3	p	○	assumption (7)
4	q		$\to$ elimination (9) on 1 and 3

We apply the same rule a second time:

1	$p \to q$	○	Assumption (7)
2	$q \to r$	○	Assumption (7)
3	p	○	assumption (7)
4	q		$\to$ elimination (9) on 1 and 3
5	r		$\to$ elimination (9) on 2 and 4

This sets up use of $\to$ introduction that we foresaw at the beginning. We'll need to apply the rule three times. First:

1	$p \to q$	○	Assumption (7)
2	$q \to r$	○	Assumption (7)
3	p	•	assumption (7)
4	q	•	$\to$ elimination (9) on 1 and 3
5	r	•	$\to$ elimination (9) on 2 and 4
6	$(p \to r)$		$\to$ introduction (8) on 3 - 5

Once again:

1	$p \to q$	○	Assumption (7)
2	$q \to r$	•	Assumption (7)
3	p	•	assumption (7)
4	q	•	$\to$ elimination (9) on 1 and 3
5	r	•	$\to$ elimination (9) on 2 and 4
6	$(p \to r)$	•	$\to$ introduction (8) on 3 - 5
7	$(q \to r) \to (p \to r)$		$\to$ introduction (8) on 2 - 6

And finally, to complete the derivation of $\emptyset$ / $(p \to q) \to ((q \to r) \to (p \to r))$:

1	$p \to q$	•	Assumption (7)
2	$q \to r$	•	Assumption (7)
3	p	•	assumption (7)
4	q	•	$\to$ elimination (9) on 1 and 3
5	r	•	$\to$ elimination (9) on 2 and 4
6	$(p \to r)$	•	$\to$ introduction (8) on 3 - 5
7	$(q \to r) \to (p \to r)$	•	$\to$ introduction (8) on 2 - 6
8	$(p \to q) \to ((q \to r) \to (p \to r))$		$\to$ introduction (8) on 1 - 7

You should verify that all three clauses of Definition (6) are indeed satisfied.

Before considering rules for other connectives, we pause for some *frequently asked questions* (FAQs) about assumptions.

> FAQ 1: Can I *really* assume any formula I want?

Yup. It's your call. Of course, such freedom doesn't make every choice of assumption wise. Some choices are smarter than others. For example, if you are trying to derive a conditional, it is always a good strategy to assume the left hand side and work to derive the right hand side.

> FAQ 2: Isn't that cheating?

No. We keep track of the assumptions with our marks. What's derived on a given line depends on previous assumptions still marked with $\circ$. It's just credit card mentality: the $\circ$ indicates a charge while $\bullet$ shows you've paid it.

> FAQ 3: I still think that $\to$ introduction is cheating since it eliminates an assumption.

It's not cheating because the rule makes visible (as the left hand side of a conditional) what used to be an assumption. Introducing $\varphi \to \psi$ into the derivation is like saying: "To derive ψ I have relied on the assumption φ."

You don't have to take our word for all this. In the next chapter we shall *prove* that any argument derivable in our system really is valid. We'll also prove the converse, namely, that every valid argument is derivable. But we're getting ahead of ourselves. First we must become acquainted with the remaining rules for extending a derivation $\boldsymbol{D}$.

> FAQ 4: When trying to prove a conditional, it is usually a good idea to assume the left hand side?

Not only is this idea *usually* good; it is *almost always* good. Only in pretty trivial situations — e.g., trying to prove $p \rightarrow q$ from $\neg\neg(p \rightarrow q)$ (see below) — should you contemplate a different strategy.

(16) EXERCISE: Show that the following arguments are derivable.

(a) $p \ / \ q \rightarrow p$

(b) $\emptyset \ / \ q \rightarrow q$

(c) $\{p \rightarrow q, q \rightarrow r\} \ / \ p \rightarrow r$

6.3.4 Conjunction rules

The rules for conditionals took a long time to explain but things will be simpler for the other connectives. Here are the rules for conjunctions.

(17) CONJUNCTION INTRODUCTION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula θ and a line with the formula ψ (in either order). Suppose also that neither of these lines bears the mark $\bullet$. Then you may append the line "$(\theta \wedge \psi)$ " to the end of $\boldsymbol{D}$.

(18) FIRST CONJUNCTION ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula $(\theta \wedge \psi)$ and that this line is not marked with $\bullet$. Then you may append the line "θ " to the end of $\boldsymbol{D}$.

(19) SECOND CONJUNCTION ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula $(\theta \wedge \psi)$ and that this line is not marked with •. Then you may append the line "ψ " to the end of $\boldsymbol{D}$.

To illustrate, let us establish the derivability of $p \wedge q \ / \ q \wedge p$. The first step is to state the premise.

1	$p \wedge q$	○	Assumption (7)

Then we apply the two elimination rules:

1	$p \wedge q$	○	Assumption (7)
2	p		$\wedge$ elimination (18)
3	q		$\wedge$ elimination (19)

Then we finish up with the introduction rule:

1	$p \wedge q$	○	Assumption (7)
2	p		$\wedge$ elimination (18)
3	q		$\wedge$ elimination (19)
4	$q \wedge p$		$\wedge$ introduction (17)

The preceding column of lines is a derivation of $q \wedge p$ from $\{p \wedge q\}$, as you can verify by consulting Definition (6).

Now for a fancier example, involving both conjunctions and conditionals. We'll provide a derivation for $p \rightarrow (q \rightarrow r) \ / \ (p \wedge q) \rightarrow r$. Since we're aiming for a conditional, we'll start by assuming its left hand side, after recording the premise in the first line.

1	$p \rightarrow (q \rightarrow r)$	○	Assumption (7)
2	$p \wedge q$	○	Assumption (7)

Then we break up the conjunction using our two elimination rules.

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	$p \wedge q$	$\circ$	Assumption (7)
3	p		$\wedge$ elimination (18)
4	q		$\wedge$ elimination (19)

The conditional elimination rule may now be used twice.

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	$p \wedge q$	$\circ$	Assumption (7)
3	p		$\wedge$ elimination (18)
4	q		$\wedge$ elimination (19)
5	$q \rightarrow r$		$\rightarrow$ elimination (9)
6	r		$\rightarrow$ elimination (9)

Conditional introduction then suffices to arrive at the desired conclusion.

1	$p \rightarrow (q \rightarrow r)$	$\circ$	Assumption (7)
2	$p \wedge q$	$\bullet$	Assumption (7)
3	p	$\bullet$	$\wedge$ elimination (18)
4	q	$\bullet$	$\wedge$ elimination (19)
5	$q \rightarrow r$	$\bullet$	$\rightarrow$ elimination (9)
6	r	$\bullet$	$\rightarrow$ elimination (9)
7	$(p \wedge q) \rightarrow r$		$\rightarrow$ introduction (8)

We're left with a derivation that ends with $(p \wedge q) \rightarrow r$ not marked by $\bullet$, and whose only line marked by $\circ$ is the argument's premise. Thus, we've derived $p \rightarrow (q \rightarrow r)$ / $(p \wedge q) \rightarrow r$

(20) EXERCISE: Derive the following arguments.

(a) $(p \wedge q) \rightarrow r$ / $p \rightarrow (q \rightarrow r)$

(b) $p \rightarrow q$ / $(p \wedge r) \rightarrow (r \wedge q)$

6.3.5 Interlude: Reiteration

If you paid close attention back in Section 3.9, you remember that a formula of form $\theta \wedge \psi$ need not have distinct conjuncts. It might be the case that θ and ψ denote the same thing, as in $\neg q \wedge \neg q$. Thus, in our conjunction rules θ and ψ do not have to represent distinct formulas. Application of $\wedge$-introduction (17) therefore allows us to derive the argument $p \;/\; p \wedge p$ as follows.

1	p	$\circ$	Assumption (7)
2	$p \wedge p$		$\wedge$ introduction (17)

This little derivation is more than a curiosity. It is sometimes useful to repeat a formula that appeared earlier in a derivation, and that is not marked by $\bullet$. The foregoing use of $\wedge$-introduction, followed by an application of $\wedge$-elimination, allows us to do so. The process can be pictured as follows.

	. . . various lines . . .		
n	θ	$\star$	the formula you want to repeat later
	. . . more lines . . .		
$n+m$	$\theta \wedge \theta$		$\wedge$ introduction (17)
$n+m+1$	θ		$\wedge$ elimination (18)

In the foregoing sketch, $\star$ represents either $\circ$ or the blank mark. You can see that this device can be used at any point in a derivation. We shall accordingly add to our basic stock of rules a *derived rule* for "reiterating" a formula. The derived rule can always be avoided by application of our conjunction rules. But a derivation may be shorter and clearer if we avail ourselves of the shortcut. It is formulated as follows.

(21) REITERATION RULE (DERIVED): Suppose that $\boldsymbol{D}$ contains a line with the formula θ and that this line is not marked with $\bullet$. Then you may append the line "θ " to the end of $\boldsymbol{D}$.

Let's take our new rule for a drive. We'll derive the argument $\{(p \to q) \to r\} \;/\; (q \to r)$. The premise and two assumptions start things off.

1	$(p \to q) \to r$	○	Assumption (7)
2	q	○	Assumption (7)
3	p	○	Assumption (7)

Now we reiterate q from line 2.

1	$(p \to q) \to r$	○	Assumption (7)
2	q	○	Assumption (7)
3	p	○	Assumption (7)
4	q		Reiteration (21)

$\to$ introduction now applies to 3 and 4.

1	$(p \to q) \to r$	○	Assumption (7)
2	q	○	Assumption (7)
3	p	●	Assumption (7)
4	q	●	Reiteration (21)
5	$p \to q$		$\to$ introduction (8)

From 1 and 5, r follows by an application of $\to$ elimination.

1	$(p \to q) \to r$	○	Assumption (7)
2	q	○	Assumption (7)
3	p	●	Assumption (7)
4	q	●	Reiteration (21)
5	$p \to q$		$\to$ introduction (8)
6	r		$\to$ elimination (9)

Finally, $q \to r$ pops out of $\to$ introduction applied to 2 and 6.

1	$(p \to q) \to r$	○	Assumption (7)
2	q	●	Assumption (7)
3	p	●	Assumption (7)
4	q	●	Reiteration (21)
5	$p \to q$	●	$\to$ introduction (8)
6	r	●	$\to$ elimination (9)
7	$q \to r$		$\to$ introduction (8)

Isn't that clever?

(22) EXERCISE:

(a) Rewrite the derivation of $\{(p \to q) \to r\}$ / $(q \to r)$ offered above, but this time without the use of the reiteration rule (21).

(b) Derive the argument $\emptyset$ / $(((q \wedge s) \to r) \to p) \to (r \to p)$

6.3.6 Disjunction rules

Back to the future. Here are the rules for disjunction.

(23) FIRST DISJUNCTION INTRODUCTION RULE: Suppose that ***D*** contains a line with the formula θ and that this line is not marked with •. Then for any formula ψ whatsoever, you may append the line "$(\theta \vee \psi)$ " to the end of ***D***.

(24) SECOND DISJUNCTION INTRODUCTION RULE: Suppose that ***D*** contains a line with the formula θ and that this line is not marked with •. Then for any formula ψ whatsoever, you may append the line "$(\psi \vee \theta)$ " to the end of ***D***.

(25) DISJUNCTION ELIMINATION RULE: Suppose that ***D*** contains lines with the formulas $(\theta \vee \psi)$, $(\theta \to \chi)$, and $(\psi \to \chi)$. (The three lines may occur in ***D*** in any order.) Suppose also that none of these lines bears the mark •. Then you may append the line "χ " to the end of ***D***.

For a simple example, let's derive $p \vee q$ / $q \vee p$. We start out:

1	$(p \vee q)$	○	Assumption (7)
2	p	○	Assumption (7)
3	$q \vee p$		$\vee$ introduction (24)

Now we use $\to$-introduction to reach:

1	$(p \vee q)$	○	Assumption (7)
2	p	•	Assumption (7)
3	$q \vee p$	•	$\vee$ introduction (24)
4	$p \rightarrow (q \vee p)$		$\rightarrow$ introduction(8)

The same process used for 2-4 is now employed to obtain the other side:

1	$(p \vee q)$	○	Assumption (7)
2	p	•	Assumption (7)
3	$q \vee p$	•	$\vee$ introduction (24)
4	$p \rightarrow (q \vee p)$		$\rightarrow$ introduction(8)
5	q	•	Assumption (7)
6	$q \vee p$	•	$\vee$ introduction (23)
7	$q \rightarrow (q \vee p)$		$\rightarrow$ introduction(8)

We then finish up with an application of $\vee$-elimination, (25) on lines 1, 4 and 7. This yields:

1	$(p \vee q)$	○	Assumption (7)
2	p	•	Assumption (7)
3	$q \vee p$	•	$\vee$ introduction (24)
4	$p \rightarrow (q \vee p)$		$\rightarrow$ introduction(8)
5	q	•	Assumption (7)
6	$q \vee p$	•	$\vee$ introduction (23)
7	$q \rightarrow (q \vee p)$		$\rightarrow$ introduction(8)
8	$(q \vee p)$		$\vee$-introduction (25)

An argument which is useful for many purposes is $(p \wedge (q \vee r))$ / $(p \wedge q) \vee (p \wedge r)$. A derivation of it starts as follows.

1	$p \wedge (q \vee r)$	○	Assumption (7)
2	$q \vee r$		$\wedge$ elimination (19)
3	q	○	Assumption (7)
4	p		$\wedge$ elimination (18)
5	$p \wedge q$		$\wedge$ introduction(17)
6	$(p \wedge q) \vee (p \wedge r)$		$\vee$-introduction (25)

An application of $\rightarrow$-introduction (8) to lines 3 - 6 then yields:

1	$p \wedge (q \vee r)$	$\circ$	Assumption (7)
2	$q \vee r$		$\wedge$ elimination (19)
3	q	$\bullet$	Assumption (7)
4	p	$\bullet$	$\wedge$ elimination (18)
5	$p \wedge q$	$\bullet$	$\wedge$ introduction(17)
6	$(p \wedge q) \vee (p \wedge r)$	$\bullet$	$\vee$-introduction(25)
7	$q \rightarrow [(p \wedge q) \vee (p \wedge r)]$		$\rightarrow$-introduction (8)

Now we repeat the whole business starting from line 3 but using r in place of q. Details:

1	$p \wedge (q \vee r)$	$\circ$	Assumption (7)
2	$q \vee r$		$\wedge$ elimination (19)
3	q	$\bullet$	Assumption (7)
4	p	$\bullet$	$\wedge$ elimination (18)
5	$p \wedge q$	$\bullet$	$\wedge$ introduction(17)
6	$(p \wedge q) \vee (p \wedge r)$	$\bullet$	$\vee$-introduction(25)
7	$q \rightarrow [(p \wedge q) \vee (p \wedge r)]$		$\rightarrow$-introduction (8)
8	r	$\bullet$	Assumption (7)
9	p	$\bullet$	$\wedge$ elimination (18)
10	$p \wedge r$	$\bullet$	$\wedge$ introduction(17)
11	$(p \wedge q) \vee (p \wedge r)$	$\bullet$	$\vee$-introduction(24)
12	$r \rightarrow [(p \wedge q) \vee (p \wedge r)]$		$\rightarrow$-introduction (8)

Lines 2, 7, and 12 then set up an application of $\vee$-elimination (25). We end up with:

1	$p \wedge (q \vee r)$	$\circ$	Assumption (7)
2	$q \vee r$		$\wedge$ elimination (19)
3	q	$\bullet$	Assumption (7)
4	p	$\bullet$	$\wedge$ elimination (18)
5	$p \wedge q$	$\bullet$	$\wedge$ introduction(17)
6	$(p \wedge q) \vee (p \wedge r)$	$\bullet$	$\vee$-introduction(25)
7	$q \rightarrow [(p \wedge q) \vee (p \wedge r)]$		$\rightarrow$-introduction (8)
8	r	$\bullet$	Assumption (7)
9	p	$\bullet$	$\wedge$ elimination (18)
10	$p \wedge r$	$\bullet$	$\wedge$ introduction(17)
11	$(p \wedge q) \vee (p \wedge r)$	$\bullet$	$\vee$-introduction(24)
12	$r \rightarrow [(p \wedge q) \vee (p \wedge r)]$		$\rightarrow$-introduction (8)
13	$(p \wedge q) \vee (p \wedge r)$		$\vee$-elimination (25)

Notice that in this last derivation the formula $((p \wedge q) \vee (p \wedge r))$ appears *three* times. This is not a mistake or inefficiency. The formula appears each time in a different role. The first time it is derived from the assumptions on lines 1 and 3, the second time from the assumptions on lines 1 and 8, and both of these are preliminary steps to deriving it on line 13 from line 1 alone.

(26) EXERCISE: Derive the following arguments.

(a) $(p \rightarrow r)$ / $((p \vee q) \rightarrow (q \vee r))$

(b) $((p \wedge q) \vee (p \wedge r))$ / $(p \wedge (q \vee r))$

(c) $r \vee (p \rightarrow q)$ / $p \rightarrow (q \vee r)$

6.3.7 Negation rules

Our rules for negation are as follows.

(27) NEGATION INTRODUCTION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula $\theta \rightarrow (\psi \wedge \neg\psi)$ not marked by $\bullet$. Then you may append the line "$\neg\theta$ " to the end of $\boldsymbol{D}$.

(28) NEGATION ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula $\neg\theta \rightarrow (\psi \wedge \neg\psi)$ not marked by $\bullet$. Then you may append the line "θ " to the end of $\boldsymbol{D}$.

As an example of the derivations that can be carried out with the negation rules, let us show that "double negation elimination" can be derived: $\neg\neg p$ / p. We start as follows.

1	$\neg\neg p$	$\circ$	Assumption (7)
2	$\neg p$	$\circ$	Assumption (7)
3	$\neg\neg p$		Reiteration (21)
4	$\neg p \wedge \neg\neg p$		$\wedge$ introduction (17)

Applying $\rightarrow$ introduction to lines 2 and 4 yields:

1	$\neg\neg p$	$\circ$	Assumption (7)
2	$\neg p$	$\bullet$	Assumption (7)
3	$\neg\neg p$	$\bullet$	Reiteration (21)
4	$\neg p \wedge \neg\neg p$	$\bullet$	$\wedge$ introduction (17)
5	$\neg p \rightarrow (\neg p \wedge \neg\neg p)$		$\rightarrow$ introduction (8)

And now we finish up with $\neg$ elimination (28), applying it to line 5.

1	$\neg\neg p$	$\circ$	Assumption (7)
2	$\neg p$	$\bullet$	Assumption (7)
3	$\neg\neg p$	$\bullet$	Reiteration (21)
4	$\neg p \wedge \neg\neg p$	$\bullet$	$\wedge$ introduction (17)
5	$\neg p \rightarrow (\neg p \wedge \neg\neg p)$		$\rightarrow$ introduction (8)
6	p		$\neg$ elimination (28)

The proof of the converse argument p / $\neg\neg p$ is very similar. It starts off:

1	p	$\circ$	Assumption (7)
2	$\neg p$	$\circ$	Assumption (7)
3	p		Reiteration (21)
4	$p \wedge \neg p$		$\wedge$ introduction (17)

Applying → introduction to lines 2 and 4 yields:

1	p	◦	Assumption (7)
2	$\neg p$	•	Assumption (7)
3	p	•	Reiteration (21)
4	$p \wedge \neg p$	•	∧ introduction (17)
5	$\neg p \rightarrow (p \wedge \neg p)$		→ introduction (8)

This time we finish up with ¬ introduction (27), applying it to line 5.

1	p	◦	Assumption (7)
2	$\neg p$	•	Assumption (7)
3	p	•	Reiteration (21)
4	$p \wedge \neg p$	•	∧ introduction (17)
5	$\neg p \rightarrow (p \wedge \neg p)$		→ introduction (8)
6	$\neg\neg p$		¬ introduction (27)

Nothing in these derivations depends on the fact that p is atomic. We could have substituted any other formula φ for all the occurrences of p and ended up with legal derivations. The new derivations would derive $\varphi \ / \ \neg\neg\varphi$ and $\neg\neg\varphi \ / \ \varphi$, respectively. Also note that the derivations do not rely on their first lines being assumptions. All that matters is that the lines are not marked by •. (If they were marked by • then reiteration could not be used at the third line.) We see, therefore, that if any line of a derivation has φ as formula and is not marked by •, we may extend the derivation by repeating the same line but with $\neg\neg\varphi$ in place of φ, and likewise with φ and $\neg\neg\varphi$ switched. It is thus possible to shorten many derivations with the following derived rule.

(29) DOUBLE NEGATION RULES (DERIVED):

(a) Suppose that $\boldsymbol{D}$ contains a line with the formula θ and that this line is not marked with •. Then you may append the line "$\neg\neg\theta$ " to the end of $\boldsymbol{D}$.

(b) Suppose that $\boldsymbol{D}$ contains a line with the formula $\neg\neg\theta$ and that this line is not marked with •. Then you may append the line "θ " to the end of $\boldsymbol{D}$.

Using (29), we have the following brief derivation of $p \wedge \neg\neg q \ / \ p \wedge q$.

1	$p \wedge \neg\neg q$	$\circ$	Assumption (7)
2	$\neg\neg q$		$\wedge$ elimination (19)
3	q		Double Negation $(29)b$
4	p		$\wedge$ elimination $(18), 1$
5	$p \wedge q$		$\wedge$ introduction $(17), 3, 4$

Without rule (29), $p \wedge \neg\neg q \ / \ p \wedge q$ would still be derivable but we would need to insert a copy of the earlier derivation for $\neg\neg p \ / \ p$ (or implement some other strategy).

Now let's derive $\neg(p \vee q) \ / \ \neg p \wedge \neg q$, one of many important relations between negation, conjunction and disjunction. Our strategy will be to derive each of $\neg p$ and $\neg q$, and our strategy for *that* will be to assume each of p and q and try to reach a contradiction. We start as follows.

1	$\neg(p \vee q)$	$\circ$	Assumption (7)
2	p	$\circ$	Assumption (7)
3	$p \vee q$		$\vee$ introduction(23)
4	$\neg(p \vee q)$		Reiteration (21)
5	$(p \vee q) \wedge \neg(p \vee q)$		$\wedge$ introduction $(17), 3, 4$

Now using $\rightarrow$ introduction and then $\neg$ introduction, we get:

1	$\neg(p \vee q)$	$\circ$	Assumption (7)
2	p	$\bullet$	Assumption (7)
3	$p \vee q$	$\bullet$	$\vee$ introduction(23)
4	$\neg(p \vee q)$	$\bullet$	Reiteration (21)
5	$(p \vee q) \wedge \neg(p \vee q)$	$\bullet$	$\wedge$ introduction $(17), 3, 4$
6	$p \rightarrow [(p \vee q) \wedge \neg(p \vee q)]$		$\rightarrow$ introduction $(8), 2, 5$
7	$\neg p$		$\neg$ introduction $(27), 6$

Next we use the same argument symmetrically with q to obtain $\neg q$, and then use $\wedge$ introduction to complete the derivation. It all looks like this:

1	$\neg(p \vee q)$	$\circ$	Assumption (7)
2	p	$\bullet$	Assumption (7)
3	$p \vee q$	$\bullet$	$\vee$ introduction(23)
4	$\neg(p \vee q)$	$\bullet$	Reiteration (21)
5	$(p \vee q) \wedge \neg(p \vee q)$	$\bullet$	$\wedge$ introduction (17), 3, 4
6	$p \rightarrow [(p \vee q) \wedge \neg(p \vee q)]$		$\rightarrow$ introduction (8), 2, 5
7	$\neg p$		$\neg$ introduction (27), 6
8	q	$\bullet$	Assumption (7)
9	$p \vee q$	$\bullet$	$\vee$ introduction(23)
10	$\neg(p \vee q)$	$\bullet$	Reiteration (21)
11	$(p \vee q) \wedge \neg(p \vee q)$	$\bullet$	$\wedge$ introduction (17), 9, 10
12	$q \rightarrow [(p \vee q) \wedge \neg(p \vee q)]$		$\rightarrow$ introduction (8), 8, 11
13	$\neg q$		$\neg$ introduction (27), 12
14	$\neg p \wedge \neg q$		$\wedge$ introduction (17)

The argument $\neg(p \vee q)$ / $\neg p \wedge \neg q$ was first explicitly noted by the English logician DeMorgan in the 19th century [24]. In the foregoing derivation, no use was made of the fact that p and q are atomic. The same derivation would go through if p were replaced by any formula θ and q by any other formula ψ. We are therefore entitled to write a new derived rule, as follows.

(30) DEMORGAN (DERIVED): Suppose that $\boldsymbol{D}$ contains a line with the formula $\neg(\theta \vee \psi)$ and that this line is not marked with $\bullet$. Then you may append the line "$\neg\theta \wedge \neg\psi$ " to the end of $\boldsymbol{D}$.

There are two other derived rules involving negations that are worth presenting. Suppose that you're working on a derivation that contains an assumption φ followed by lines with θ and $\neg\theta$ as formulas. We can picture the situation like this:

n	φ	$\circ$	Assumption (7)
	. . . various lines . . .		
$n+k$	θ		
	. . . more lines . . .		
$n+k+\ell$	$\neg\theta$		

You see that any such derivation can be extended via $\wedge$ introduction and $\rightarrow$ introduction as follows.

n	φ	$\bullet$	Assumption (7)
	. . . various lines . . .		
$n+k$	θ	$\bullet$	
	. . . more lines . . .		
$n+k+\ell$	$\neg\theta$	$\bullet$	
$n+k+\ell+1$	$\theta \wedge \neg\theta$	$\bullet$	$\wedge$ introduction (17)
$n+k+\ell+2$	$\varphi \rightarrow (\theta \wedge \neg\theta)$		$\rightarrow$ introduction (8)

With line $n + k + \ell + 2$ in hand, $\neg\varphi$ can now be added via $\neg$ introduction (27). The whole derivation looks like this:

n	φ	$\bullet$	Assumption (7)
	. . . various lines . . .		
$n+k$	θ	$\bullet$	
	. . . more lines . . .		
$n+k+\ell$	$\neg\theta$	$\bullet$	
$n+k+\ell+1$	$\theta \wedge \neg\theta$	$\bullet$	$\wedge$ introduction (17)
$n+k+\ell+2$	$\varphi \rightarrow (\theta \wedge \neg\theta)$		$\rightarrow$ introduction (8)
$n+k+\ell+3$	$\neg\varphi$		$\neg$ introduction (27)

It doesn't matter whether θ or $\neg\theta$ comes first in the foregoing derivation (lines $n + k$ and $n + k + \ell$ could be switched). Also, it doesn't matter whether we end up with the conjunction on line $n + k + \ell + 1$ via conjunction introduction or through some other route. So we write a new derived rule as follows.

(31) NEGATION INTRODUCTION (DERIVED): Suppose that the last line in $\boldsymbol{D}$ marked by $\circ$ has φ as formula. (Don't use this rule if no line in $\boldsymbol{D}$ is marked by $\circ$.) Suppose also that *either* (a) there are two subsequent lines in $\boldsymbol{D}$, neither marked by $\bullet$, and containing the formulas θ and $\neg\theta$, *or* (b) there is a subsequent line in $\boldsymbol{D}$ unmarked by $\bullet$ containing either $\theta \wedge \neg\theta$ or $\neg\theta \wedge \theta$. Then you may do the following. *First*, from "φ $\circ$" to the end of $\boldsymbol{D}$, change the mark of every line to $\bullet$ (if the mark is not $\bullet$ already). *Next*, append the line "$\neg\varphi$ " to the end of $\boldsymbol{D}$.

If you examine the derivation sketch above, you'll see that the roles of φ and $\neg\varphi$ can be reversed. In this case, the last line involves φ, and is justified by $\neg$ elimination (28) instead of $\neg$ introduction (27). We can therefore write another derived rule, symmetrical to (31).

(32) NEGATION ELIMINATION (DERIVED): Suppose that the last line in $\boldsymbol{D}$ marked by $\circ$ has $\neg\varphi$ as formula. (Don't use this rule if no line in $\boldsymbol{D}$ is marked by $\circ$.) Suppose also that *either* (a) there are two subsequent lines in $\boldsymbol{D}$, neither marked by $\bullet$, and containing formulas θ and $\neg\theta$, *or* (b) there is a subsequent line in $\boldsymbol{D}$ unmarked by $\bullet$ containing either $\theta \wedge \neg\theta$ or $\neg\theta \wedge \theta$. Then you may do the following. *First*, from "$\neg\varphi$ $\circ$" to the end of $\boldsymbol{D}$, change the mark of every line to $\bullet$ (if the mark is not $\bullet$ already). *Next*, append the line "φ " to the end of $\boldsymbol{D}$.

Let's put the derived $\neg$ introduction rule (31) to use by deriving $p \rightarrow q$ / $\neg q \rightarrow \neg p$. This argument is traditionally called *contraposition*. The derivation gets going as follows.

1	$p \rightarrow q$	$\circ$	Assumption (7)
2	$\neg q$	$\circ$	Assumption (7)

We've thus assumed the left hand side of our goal, and we now want to derive the right hand side. Since the principal connective of the right hand side is $\neg$, derived $\neg$ introduction may be helpful. We therefore assume the formula without the negation and look for a contradiction.

1	$p \rightarrow q$	$\circ$	Assumption (7)
2	$\neg q$	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)
4	q		$\rightarrow$ elimination $(9), 1, 3$
5	$\neg q$		Reiteration $(21), 2$

Now we can use our derived $\neg$ introduction rule to obtain:

1	$p \rightarrow q$	$\circ$	Assumption (7)
2	$\neg q$	$\circ$	Assumption (7)
3	p	$\bullet$	Assumption (7)
4	q	$\bullet$	$\rightarrow$ elimination $(9), 1, 3$
5	$\neg q$	$\bullet$	Reiteration $(21), 2$
6	$\neg p$		Derived $\neg$ elimination$(32), 3, 4, 5$

We finish with $\rightarrow$ introduction:

1	$p \rightarrow q$	$\circ$	Assumption (7)
2	$\neg q$	$\bullet$	Assumption (7)
3	p	$\bullet$	Assumption (7)
4	q	$\bullet$	$\rightarrow$ elimination $(9), 1, 3$
5	$\neg q$	$\bullet$	Reiteration $(21), 2$
6	$\neg p$	$\bullet$	derived $\neg$ elimination$(32), 3, 4, 5$
7	$\neg q \rightarrow \neg p$		$\rightarrow$ introduction$(8), 2, 6$

There's one more derived rule involving negation that we'd like to bring to your attention. It will be used later, and anyway it's fun to think about.

(33) CONTRADICTION RULE (DERIVED): Suppose that ***D*** contains a line with a formula of the form $\varphi \wedge \neg\varphi$, not marked by $\bullet$. Then you may append *any* line marked with the blank.

To see that (33) is justified, consider a derivation that ends with $\varphi \wedge \neg\varphi$, not marked by $\bullet$. Then it can be continued as pictured here (where $\star$ is either $\circ$ or blank).

n	$\varphi \wedge \neg\varphi$	$\star$	
$n+1$	$\neg\psi$	$\circ$	Assumption (7)
$n+2$	$\varphi \wedge \neg\varphi$		Reiteration (21)

In the foregoing, ψ can be any formula you choose. The assumption of $\neg\psi$ can then be cancelled by introduction of a conditional followed by the Negation Elimination Rule (28). The whole thing is pictured as follows.

n	$\varphi \wedge \neg\varphi$	$\star$	
$n+1$	$\neg\psi$	$\bullet$	Assumption (7)
$n+2$	$\varphi \wedge \neg\varphi$	$\bullet$	Reiteration (21)
$n+3$	$\neg\psi \rightarrow (\varphi \wedge \neg\varphi)$		$\rightarrow$ Introduction (8)
$n+4$	ψ		$\neg$ elimination (28)

In Section 5.3.2 we considered the fact that $p \wedge \neg p \models \psi$ for all $\psi \in \mathcal{L}$. To make this feature of logic more palatable, we provided an informal derivation of (arbitrary) ψ from $p \wedge \neg p$. Now we possess an "official" derivation of the same fact.

(34) EXERCISE: Demonstrate that the following means of extending $\boldsymbol{D}$ is a derived rule (that is, its use can always be replaced by recourse to the basic rules).

> NEGATION INTRODUCTION (DERIVED): Suppose that the last line in $\boldsymbol{D}$ marked with $\circ$ has φ as formula. (Don't use this rule if no line in $\boldsymbol{D}$ is marked by $\circ$.) Suppose also that the last line of $\boldsymbol{D}$ has $\neg\varphi$. Then you may do the following. *First*, from "φ $\circ$" to the end of $\boldsymbol{D}$, change the mark of every line to $\bullet$ (if the mark is not $\bullet$ already). *Next*, append the line "$\neg\varphi$ " to the end of $\boldsymbol{D}$.

There is a symmetrical derived rule for eliminating $\neg$. Can you formulate it?

(35) EXERCISE: Provide derivations for the following arguments. Feel free to use the derived rules that we have established.

(a) $q \;/\; \neg(p \wedge \neg p)$

(b) $p \wedge \neg q \;/\; \neg(p \rightarrow q)$

(c) $\neg p \vee \neg q \;/\; \neg(p \wedge q)$

(d) $p \wedge \neg p \;/\; \neg q$

6.3.8 Biconditional rules

Finally, we arrive at the last connective in $\mathcal{L}$, the biconditional. Its rules are as follows.

(36) BICONDITIONAL INTRODUCTION RULE: Suppose that ***D*** contains a line with the formula $(\theta \rightarrow \psi)$ and a line with the formula $(\psi \rightarrow \theta)$ (in either order). Suppose also that neither of these lines bears the mark $\bullet$. Then you may append the line "$\theta \leftrightarrow \psi$ " to the end of ***D***.

(37) FIRST BICONDITIONAL ELIMINATION RULE: Suppose that ***D*** contains a line with the formula θ and a line with the formula $(\theta \leftrightarrow \psi)$ (in either order). Suppose also that neither of these lines bears the mark $\bullet$. Then you may append the line "ψ " to the end of ***D***.

(38) SECOND BICONDITIONAL ELIMINATION RULE: Suppose that ***D*** contains a line with the formula ψ and a line with the formula $(\theta \leftrightarrow \psi)$ (in either order). Suppose also that neither of these lines bears the mark $\bullet$. Then you may append the line "θ " to the end of ***D***.

Let's see how to use these rules to derive the argument with premises $\{p \leftrightarrow q, q \leftrightarrow r\}$ and conclusion $p \leftrightarrow r$. We start like this:

1	$p \leftrightarrow q$	$\circ$	Assumption (7)
2	$q \leftrightarrow r$	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)
4	q		$\leftrightarrow$ elimination $(37), 1, 3$
5	r		$\leftrightarrow$ elimination $(37), 2, 4$

The assumption at line 3 can now be canceled to get the first conditional we need:

1	$p \leftrightarrow q$	○	Assumption (7)
2	$q \leftrightarrow r$	○	Assumption (7)
3	p	●	Assumption (7)
4	q	●	$\leftrightarrow$ elimination (37), 1, 3
5	r	●	$\leftrightarrow$ elimination (37), 2, 4
6	$p \rightarrow r$		$\rightarrow$ introduction (8), 3, 5

Obtaining the converse conditional $r \rightarrow p$ is achieved in the same way. Once it is in hand, we finish with $\leftrightarrow$ introduction (36). The whole thing looks like this:

1	$p \leftrightarrow q$	○	Assumption (7)
2	$q \leftrightarrow r$	○	Assumption (7)
3	p	●	Assumption (7)
4	q	●	$\leftrightarrow$ elimination (37), 1, 3
5	r	●	$\leftrightarrow$ elimination (37), 2, 4
6	$p \rightarrow r$		$\rightarrow$ introduction (8), 3, 5
7	r	●	Assumption (7)
8	q	●	$\leftrightarrow$ elimination (38), 2, 7
9	p	●	$\leftrightarrow$ elimination (38), 1, 8
10	$r \rightarrow p$		$\rightarrow$ introduction (8), 7, 10
11	$p \leftrightarrow r$		$\leftrightarrow$ introduction (36), 6, 10

That was easy. For something more challenging, let's derive $p \leftrightarrow q$ / $(p \vee r) \leftrightarrow (r \vee q)$. Things start off with:

1	$p \leftrightarrow q$	○	Assumption (7)
2	$p \vee r$	○	Assumption (7)
3	p	○	Assumption (7)
4	q		$\leftrightarrow$ elimination (37), 1, 3
5	$r \vee q$		$\vee$ introduction (24), 4

Now we add the first needed conditional, and cancel everything from line 3.

1	$p \leftrightarrow q$	○	Assumption (7)
2	$p \vee r$	○	Assumption (7)
3	p	•	Assumption (7)
4	q	•	$\leftrightarrow$ elimination $(37), 1, 3$
5	$r \vee q$	•	$\vee$ introduction $(24), 4$
6	$p \rightarrow (r \vee q)$		$\rightarrow$ introduction $(8), 3, 5$

Next, let us assume r, and add steps similar to (3) - (6).

1	$p \leftrightarrow q$	○	Assumption (7)
2	$p \vee r$	○	Assumption (7)
3	p	•	Assumption (7)
4	q	•	$\leftrightarrow$ elimination $(37), 1, 3$
5	$r \vee q$	•	$\vee$ introduction $(24), 4$
6	$p \rightarrow (r \vee q)$		$\rightarrow$ introduction $, 3, 5$
7	r	•	Assumption (7)
8	$r \vee q$	•	$\vee$ introduction $(23), 4$
9	$r \rightarrow (r \vee q)$		$\rightarrow$ introduction $(8), 3, 5$

Lines 2, 6, and 9 are now exploited by $\vee$ elimination, rule (25), as follows.

1	$p \leftrightarrow q$	○	Assumption (7)
2	$p \vee r$	○	Assumption (7)
3	p	•	Assumption (7)
4	q	•	$\leftrightarrow$ elimination $(37), 1, 3$
5	$r \vee q$	•	$\vee$ introduction $(24), 4$
6	$p \rightarrow (r \vee q)$		$\rightarrow$ introduction $, 3, 5$
7	r	•	Assumption (7)
8	$r \vee q$	•	$\vee$ introduction $(23), 4$
9	$r \rightarrow (r \vee q)$		$\rightarrow$ introduction $(8), 3, 5$
10	$r \vee q$		$\vee$ elimination $, 2, 6, 9$

Lines 2 and 10 suffice to obtain the first conditional needed to derive the biconditional we seek. Specifically:

1	$p \leftrightarrow q$	○	Assumption (7)
2	$p \vee r$	•	Assumption (7)
3	p	•	Assumption (7)
4	q	•	$\leftrightarrow$ elimination $(37), 1, 3$
5	$r \vee q$	•	$\vee$ introduction $(24), 4$
6	$p \rightarrow (r \vee q)$	•	$\rightarrow$ introduction $(8), 3, 5$
7	r	•	Assumption (7)
8	$r \vee q$	•	$\vee$ introduction $(23), 4$
9	$r \rightarrow (r \vee q)$	•	$\rightarrow$ introduction $(8), 3, 5$
10	$r \vee q$	•	$\vee$ elimination $, 2, 6, 9$
11	$(p \vee r) \rightarrow (r \vee q)$		$\rightarrow$ introduction $(8), 2, 10$

It remains to derive the converse conditional, setting up an application of $\leftrightarrow$ introduction (36). Our strategy is similar to what's already been produced, and brings us to line 21 in the following derivation. The final line 22 is obtained by combining two conditionals, as described in $\leftrightarrow$ introduction (36), above.

1	$p \leftrightarrow q$	○	Assumption (7)
2	$p \vee r$	•	Assumption (7)
3	p	•	Assumption (7)
4	q	•	$\leftrightarrow$ **elimination** $(37), 1, 3$
5	$r \vee q$	•	$\vee$ **introduction** $(24), 4$
6	$p \rightarrow (r \vee q)$	•	$\rightarrow$ **introduction** $(8), 3, 5$
7	r	•	Assumption (7)
8	$r \vee q$	•	$\vee$ **introduction** $(23), 4$
9	$r \rightarrow (r \vee q)$	•	$\rightarrow$ **introduction** $(8), 3, 5$
10	$r \vee q$	•	$\vee$ **elimination** $, 2, 6, 9$
11	$(p \vee r) \rightarrow (r \vee q)$		$\rightarrow$ **introduction** $(8), 2, 10$
12	$r \vee q$	•	Assumption (7)
13	r	•	Assumption (7)
14	$p \vee r$	•	$\vee$ **introduction** $(24), 13$
15	$r \rightarrow (p \vee r)$	•	$\rightarrow$ **introduction** $(8), 13, 14$
16	q	•	Assumption (7)
17	p	•	$\leftrightarrow$ **elimination** (38)
18	$p \vee r$	•	$\vee$ **introduction** $(23), 4$
19	$q \rightarrow (p \vee r)$	•	$\rightarrow$ **introduction** $(8), 16, 18$
20	$p \vee r$	•	$\vee$ **elimination** $, 12, 16, 19$
21	$(q \vee r) \rightarrow (p \vee r)$		$\rightarrow$ **introduction** $(8), 12, 20$
22	$(p \vee r) \leftrightarrow (r \vee q)$		$\leftrightarrow$ **introduction** $(36), 11, 21$

Now it's time for *you* to do some work.

(39) EXERCISE: Give derivations for the following arguments. Feel free to use derived rules.

(a) $\emptyset \;/\; (p \leftrightarrow q) \leftrightarrow [(p \rightarrow q) \wedge (q \rightarrow p)]$

(b) $p \wedge q \;/\; p \leftrightarrow q$

(c) $p \leftrightarrow q \;/\; (p \wedge q) \leftrightarrow (r \wedge q)$

(d) $\{p \leftrightarrow q, r \leftrightarrow s\} \;/\; (p \wedge r) \leftrightarrow (q \wedge s)$

6.4 Indirect Proof

Our negation rules (27) and (28) — along with their derived extensions (31) - (34) — are related to a form of argument that has traditionally been known as “indirect proof.” Indirect arguments proceed by assuming the opposite of what they seek to conclude. The assumption leads to “absurdity” in the form of a contradiction, so this kind of reasoning is also said to involve *reductio ad absurdum*. The pivotal contradiction must involve at least one formula marked with the blank, as we shall see in the illustrations to follow. Facility with the indirect strategy is essential to developing skill in finding derivations for arguments, including some arguments whose conclusions are not negations.

Let’s start by considering disjunctions. Suppose we wish to prove a formula of form $\varphi \vee \psi$. The first strategy that comes to mind is to try to prove either φ or ψ, and then apply one of the $\vee$ introduction rules (23), (24). This idea doesn’t always work, however. For example, if we want to derive the argument $\emptyset \ / \ p \vee \neg p$, it is futile to try first to derive either $\emptyset \ / \ p$ or $\emptyset \ / \ \neg p$. In the next chapter it will be seen that neither of the latter, invalid, arguments can be derived. To derive $\emptyset \ / \ p \vee \neg p$ we must rather proceed indirectly, setting things up for an application of $\neg$ elimination [in its derived form (32)]. Here is how we get started.

1	$\neg(p \vee \neg p)$	$\circ$	Assumption (7)
2	$\neg p \wedge \neg\neg p$		DeMorgan (30)

Now we can use $\neg$ elimination to finish up.

1	$\neg(p \vee \neg p)$	$\bullet$	Assumption (7)
2	$\neg p \wedge \neg\neg p$	$\bullet$	DeMorgan (30)
3	$p \vee \neg p$		Derived $\neg$ elimination $(32), 1, 2$

For another example, consider the following argument which exhibits an important relation between conditionals, disjunction, and negation: $p \rightarrow q \ / \ \neg p \vee q$. Again, we won’t make progress by trying to derive either $\neg p$ or q from the premise $p \rightarrow q$. This is because the (invalid) arguments $p \rightarrow q \ / \ \neg p$ and $p \rightarrow q \ / \ q$ are not derivable in our system, as will be seen in the next chapter. To proceed indirectly, we must assume the negation of the conclusion and then hunt for a contradiction. So we start this way:

1	$p \to q$	∘	Assumption (7)
2	$\neg(\neg p \vee q)$	∘	Assumption (7)
3	$\neg\neg p \wedge \neg q$		DeMorgan (30), 2
4	$\neg\neg p$		$\wedge$ elimination (9), 3
5	p		derived double $\neg$ rule (29)b, 4
6	q		$\to$ elimination (9)
7	$\neg q$		$\wedge$ elimination (19)

Now we finish up with the derived version of $\neg$ elimination:

1	$p \to q$	∘	Assumption (7)
2	$\neg(\neg p \vee q)$	•	Assumption (7)
3	$\neg\neg p \wedge \neg q$	•	DeMorgan (30), 2
4	$\neg\neg p$	•	$\wedge$ elimination (9), 3
5	p	•	derived double $\neg$ rule (29)b, 4
6	q	•	$\to$ elimination (9)
7	$\neg q$	•	$\wedge$ elimination (19)
8	$\neg p \vee q$		derived $\neg$ elimination (32), 2, 6, 7

In the preceding examples, we derived arguments whose conclusions have nontrivial logical structure (they both were disjunctions). Our indirect strategy also applies to cases where the conclusion is just a variable, as in the argument $\{\neg p \to q, \neg q\}$ / p. Our strategy here is to assume $\neg p$, and hunt for a contradiction. We start off this way.

1	$\neg p \to q$	∘	Assumption (7)
2	$\neg q$	∘	Assumption (7)
3	$\neg p$	∘	Assumption (7)
4	q		$\to$ elimination (9)
5	$\neg q$		Reiteration (21)

The use of Reiteration at line (5) allows us to finish up with $\neg$ elimination:

1	$\neg p \to q$	$\circ$	Assumption (7)
2	$\neg q$	$\circ$	Assumption (7)
3	$\neg p$	$\bullet$	Assumption (7)
4	q	$\bullet$	$\to$ elimination (9)
5	$\neg q$	$\bullet$	Reiteration (21)
6	p		derived $\neg$ elimination $(32), 2, 4, 5$

How do you know when you should employ the indirect strategy? Since there may be no overt clue that $\neg$ elimination will pay off, we formulate our Most Important Strategic Principle. It has two parts.

What to do when you don't know what to do:

(a) Don't panic!

(b) Assume the negation of your goal (or subgoal) and look for a contradiction.

An excellent discussion of Part (a) is already available in D. Adams [2]. We concentrate on (b).

Principle (b) may be useful at the beginning of a derivation, for example, to derive the argument $\emptyset$ / $(p \to q) \vee (q \to r)$. (You'll be asked to derive it as an exercise, below.) The advice might also serve you well in the middle of a derivation, when you're stuck. To illustrate, let us try to establish the somewhat surprising argument known as *Pierce's Law*, $\emptyset$ / $((p \to q) \to p) \to p$. Since the conclusion is a conditional, we can start by assuming the left hand side.

1	$(p \to q) \to p$	$\circ$	Assumption (7)

If we could derive $p \to q$ then we could use $\to$ elimination to get p. But we have nothing to work with to get $p \to q$. Not knowing what to do, we use Principle (b) and assume $\neg p$. With $\neg p$ we can get to work on deriving a contradiction involving at least one line marked with the blank. For the latter purpose, we assume p in view of deriving q (for an application of $\to$ introduction). Thus:

1	$(p \rightarrow q) \rightarrow p$	$\circ$	Assumption (7)
2	$\neg p$	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)

Our immediate goal is now to get q. Once again, we have no idea how to achieve this so we apply our principle again by assuming $\neg q$. This is followed by two uses of Reiteration in order to set up $\neg$ elimination.

1	$(p \rightarrow q) \rightarrow p)$	$\circ$	Assumption (7)
2	$\neg p$	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)
4	$\neg q$	$\circ$	Assumption (7)
5	$\neg p$		Reiteration $(21), 2$
6	p		Reiteration $(21), 3$

Now we can use $\neg$ elimination to get q.

1	$(p \rightarrow q) \rightarrow p)$	$\circ$	Assumption (7)
2	$\neg p$	$\circ$	Assumption (7)
3	p	$\circ$	Assumption (7)
4	$\neg q$	$\bullet$	Assumption (7)
5	$\neg p$	$\bullet$	Reiteration $(21), 2$
6	p	$\bullet$	Reiteration $(21), 3$
7	q		derived $\neg$ elimination $(32), 4, 5, 6$

We can now get to $p \rightarrow q$, thence to p.

1	$(p \rightarrow q) \rightarrow p)$	$\circ$	Assumption (7)
2	$\neg p$	$\circ$	Assumption (7)
3	p	$\bullet$	Assumption (7)
4	$\neg q$	$\bullet$	Assumption (7)
5	$\neg p$	$\bullet$	Reiteration $(21), 2$
6	p	$\bullet$	Reiteration $(21), 3$
7	q	$\bullet$	derived $\neg$ elimination $(32), 4, 5, 6$
8	$p \rightarrow q$		$\rightarrow$ introduction $(8), 3, 7$
9	p		$\rightarrow$ elimination $(9), 1, 8$

We are ready to finish, first by using $\neg$ elimination on line 2, then by using $\rightarrow$ introduction on 1 and 11. As a preliminary, we bring $\neg p$ from line 2 to line 10 via Reiteration. We'll show these steps in two stages. Here is the derivation after $\neg$ elimination.

1	$(p \rightarrow q) \rightarrow p)$	○	Assumption (7)
2	$\neg p$	•	Assumption (7)
3	p	•	Assumption (7)
4	$\neg q$	•	Assumption (7)
5	$\neg p$	•	Reiteration $(21), 2$
6	p	•	Reiteration $(21), 3$
7	q	•	derived $\neg$ elimination $(32), 4, 5, 6$
8	$p \rightarrow q$	•	$\rightarrow$ introduction $(8), 3, 7$
9	p	•	$\rightarrow$ elimination $(9), 1, 8$
10	$\neg p$	•	Reiteration $, 2$
11	p		derived $\neg$ elimination $(32), 2, 9, 10$

And here is the *coup de grace* via $\rightarrow$ introduction.

1	$(p \rightarrow q) \rightarrow p)$	•	Assumption (7)
2	$\neg p$	•	Assumption (7)
3	p	•	Assumption (7)
4	$\neg q$	•	Assumption (7)
5	$\neg p$	•	Reiteration $(21), 2$
6	p	•	Reiteration $(21), 3$
7	q	•	derived $\neg$ elimination $(32), 4, 5, 6$
8	$p \rightarrow q$	•	$\rightarrow$ introduction $(8), 3, 7$
9	p	•	$\rightarrow$ elimination $(9), 1, 8$
10	$\neg p$	•	Reiteration $, 2$
11	p	•	derived $\neg$ elimination $(32), 2, 9, 10$
12	$(p \rightarrow q) \rightarrow p) \rightarrow p$		$\rightarrow$ introduction $(8), 1, 11$

Did you follow all that? If not, you might wish to go back over the material in this section. It's the indirect strategy that's the real challenge in derivations.

(40) EXERCISE: Find derivations for the following arguments. Feel free to use derived rules.

(a) $\neg(p \wedge q)$ / $\neg p \vee \neg q$

(b) $\neg(p \rightarrow q)$ / $p \wedge \neg q$

(c) $p \rightarrow (q \vee r)$ / $r \vee (p \rightarrow q)$

(d) $\emptyset$ / $(p \rightarrow q) \vee (q \rightarrow r)$

(e) $(p \rightarrow q) \rightarrow p$ / $p \vee q$

(41) EXERCISE: Show that the $\neg$ introduction rule (27) is redundant in the sense that its use can be simulated with the other non-derived rules. In other words, show that we could suppress (27) without losing the ability to derive any argument. In your proof, make sure not to rely on any of the derived rules. (You get extra credit for this one.)

6.5 Derivation of formulas, interderivability

In discussing Pierce's Law above, we considered an argument with the empty set of premises. This is just a special case of Definition (6) but it arises often enough to deserve special recognition.

(42) DEFINITION: A *derivation of the formula* φ is a derivation with the following properties.

(a) The derivation ends with the line "φ ".

(b) No lines in the derivation are marked by $\circ$.

If φ has such a derivation, then we say that φ is *derivable*.

To avoid misunderstanding, we note that the derivation showing φ to be derivable may contain $\circ$ *during its construction*, just not when it's finished. It should be clear that φ is derivable in the sense of Definition (42) if and only if $\emptyset$ / φ is derivable in the sense of Definition (6). To illustrate, the discussion of Pierce's Law, above, informs us that $((p \to q) \to p) \to p$ is derivable.

Call formulas θ, ψ *interderivable* if both the arguments $\theta \ / \ \psi$ and $\psi \ / \ \theta$ are derivable. For example, $p \wedge q$ and $q \wedge p$ are interderivable, as you can easily check. You should also be able to verify the following fact, using the rules for conditionals and biconditionals in Sections 6.3.3 and 6.3.8.

(43) FACT: Let $\theta, \psi \in \mathcal{L}$ be given. Then θ, ψ are interderivable if and only if $\theta \leftrightarrow \psi$ is derivable.

The following corollary is almost immediate (we leave its proof to you).

(44) COROLLARY: For all $\theta, \psi, \varphi \in \mathcal{L}$, if θ and ψ are interderivable, and ψ and φ are interderivable then θ and φ are interderivable.

Of course, the corollary applies to chains of any length. For example, if θ_1 and θ_2 are interderivable, θ_2 and θ_3 are interderivable, and θ_3 and θ_4 are interderivable then θ_1 and θ_4 are interderivable. This follows by applying the corollary a first time to obtain the interderivability of θ_1 and θ_3, then a second time to obtain the interderivability of θ_1 and θ_4. Chains of any length can be treated the same way. In view of Corollary (44), we say that interderivability is a "transitive" relation.

6.6 Derivation schemas

In thinking about derivability, it often helps to write down a *derivation schema*. Such a schema is a blueprint for official derivations. It relies on Greek letters to represent arbitrary formulas, along with other notations. We saw this kind of thing when justifying the derived rule for negation elimination (32). For another illustration, let us convince ourselves of the following claim.

(45) FACT: For all $\varphi, \theta \in \mathcal{L}$, if $\varphi \leftrightarrow \theta$ is derivable then so is $\neg\varphi \leftrightarrow \neg\theta$.

The proof consists of picturing a derivation of $\varphi \leftrightarrow \theta$ that is extended to one for $\neg\varphi \leftrightarrow \neg\theta$. Here is how it might look when finished.

1	$\varphi \leftrightarrow \theta$		assumed to be derivable
2	$\neg\varphi$	•	cancelled assumption (7)
3	θ	•	cancelled assumption (7)
4	φ	•	$\leftrightarrow$ elimination $(38), 1, 3$
5	$\neg\varphi$	•	Reiteration $(21), 2$
6	$\neg\theta$	•	derived $\neg$ introduction $(31), 3, 4, 5$
7	$\neg\varphi \rightarrow \neg\theta$		$\rightarrow$ introduction $(8), 2, 6$
8	$\neg\theta$	•	cancelled assumption (7)
9	φ	•	cancelled assumption (7)
10	θ	•	$\leftrightarrow$ elimination $(37), 1, 9$
11	$\neg\theta$	•	Reiteration $(21), 8$
12	$\neg\varphi$	•	derived $\neg$ introduction $(31), 10, 11, 12$
13	$\neg\theta \rightarrow \neg\varphi$		$\rightarrow$ introduction $(8), 8, 12$
14	$\neg\varphi \leftrightarrow \neg\theta$		$\leftrightarrow$ introduction $(36), 7, 13$

Let's do one more example. Suppose you'd like to demonstrate the following fact.

(46) FACT: For all $\varphi_1, \varphi_2, \theta_1, \theta_2 \in \mathcal{L}$, if $\varphi_1 \leftrightarrow \theta_1$ and $\varphi_2 \leftrightarrow \theta_2$ are both derivable then so is $(\varphi_1 \wedge \varphi_2) \leftrightarrow (\theta_1 \wedge \theta_2)$.

Your argument for (46) would take the form of a derivation schema that might end up looking like this:

1	$\varphi_1 \leftrightarrow \theta_1$		assumed to be derivable
2	$\varphi_2 \leftrightarrow \theta_2$		also assumed to be derivable
3	$\varphi_1 \wedge \varphi_2$	•	cancelled assumption (7)
4	φ_1	•	$\wedge$ elimination$(18), 3$
5	θ_1	•	$\leftrightarrow$ elimination$(37), 1, 4$
6	φ_2	•	$\wedge$ elimination$(18)3$
7	θ_2	•	$\leftrightarrow$ elimination$(37)2, 6$
8	$\theta_1 \wedge \theta_2$		$\wedge$ introduction$(17), 5, 7$
9	$(\varphi_1 \wedge \varphi_2) \to (\theta_1 \wedge \theta_2)$		$\to$ introduction$(8), 3, 8$
10	$\theta_1 \wedge \theta_2$	•	cancelled assumption (7)
11	θ_1	•	$\wedge$ elimination$(18), 10$
12	φ_1	•	$\leftrightarrow$ elimination$(38), 1, 11$
13	θ_2	•	$\wedge$ elimination$(18)10$
14	φ_2	•	$\leftrightarrow$ elimination$(38)2, 13$
15	$\varphi_1 \wedge \varphi_2$	•	$\wedge$ introduction$(17), 12, 14$
16	$(\varphi_1 \wedge \varphi_2) \to (\theta_1 \wedge \theta_2)$		$\to$ introduction$(8), 10, 16$
17	$(\varphi_1 \wedge \varphi_2) \leftrightarrow (\theta_1 \wedge \theta_2)$		$\leftrightarrow$ introduction $(36), 9, 16$

Facts similar to (45) and (46) are recorded below. We will rely on them in the next chapter in order to exhibit fundamental properties of derivability. It would be too painful to verify them all *in extenso*. Their proofs are therefore left to you, with permission to sacrifice some explicitness for brevity.

(47) FACT: Let $\varphi_1, \varphi_2, \theta_1, \theta_2 \in \mathcal{L}$ be given. Suppose that $\varphi_1 \leftrightarrow \theta_1$ and $\varphi_2 \leftrightarrow \theta_2$ are both derivable. Then:

(a) $(\varphi_1 \vee \varphi_2) \leftrightarrow (\theta_1 \vee \theta_2)$ is derivable.

(b) $(\varphi_1 \to \varphi_2) \leftrightarrow (\theta_1 \to \theta_2)$ is derivable.

(c) $(\varphi_1 \leftrightarrow \varphi_2) \leftrightarrow (\theta_1 \leftrightarrow \theta_2)$ is derivable.

(48) FACT: Let $\psi, \chi, \gamma \in \mathcal{L}$ be given. Then the following pairs of formulas are interderivable.

(a) $\psi \to \chi$ and $\neg\psi \vee \chi$

(b) $\psi \leftrightarrow \chi$ and $(\psi \wedge \chi) \vee (\neg\psi \wedge \neg\chi)$

(c) $\neg\neg\psi$ and ψ

(d) $\neg(\chi \wedge \psi)$ and $\neg\chi \vee \neg\psi$

(e) $\neg(\chi \vee \psi)$ and $\neg\chi \wedge \neg\psi$

(f) $\psi \wedge (\chi \vee \gamma)$ and $(\psi \wedge \chi) \vee (\psi \wedge \gamma)$

(g) $(\chi \vee \gamma) \wedge \psi$ and $(\chi \wedge \psi) \vee (\gamma \wedge \psi)$

Facts (48)d,e are usually called "DeMorgan laws." [We stated something similar in (30).] Facts (48)f,g are usually called "Distribution laws." Other distribution laws can also be proved. They switch the roles of $\wedge$ and $\vee$ in (48)f. We don't bother to state them officially, however, since they won't be used later. We *will* need the following (and final) three facts.

(49) FACT: Suppose that the argument Γ / $\varphi \wedge \neg\varphi$ is derivable, for some $\varphi \in \mathcal{L}$. Then the argument Γ / $p \wedge \neg p$ is also derivable.

(50) FACT: Let $\theta_1 \cdots \theta_n, \varphi \in \mathcal{L}$ be given. Suppose that the arguments θ_i / φ are all derivable. Then the argument $\theta_1 \vee \cdots \vee \theta_n$ /φ is also derivable.

(51) FACT: Let $\theta_1 \cdots \theta_n, \varphi \in \mathcal{L}$ be given, and suppose that the zero-premise argument $\theta_1 \wedge \cdots \wedge \theta_n \rightarrow \varphi$ is derivable. Then the n-premise argument $\theta_1 \cdots \theta_n$ / φ is derivable.

6.7 Summary of rules

We bring together all the rules presented in this chapter.

(52) ASSUMPTION RULE: For any formula φ whatsoever you may append the line "φ ∘" to the end of $\boldsymbol{D}$.

(53) CONDITIONAL INTRODUCTION RULE: Suppose that $\boldsymbol{D}$ ends with a line whose formula is ψ. Suppose also that this line is not marked by •. Let

"θ ∘" be the last line in $\boldsymbol{D}$ marked with ∘. (If there is no line in $\boldsymbol{D}$ marked with ∘ then you cannot use this rule.) Then you may do the following. *First*, from "θ ∘" to the end of $\boldsymbol{D}$, change the mark of every line to • (if the mark is not • already). *Next*, append the line "$\theta \to \psi$ " to the end of $\boldsymbol{D}$.

(54) CONDITIONAL ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula θ and a line with the formula $(\theta \to \psi)$ (in either order). Suppose also that neither of these lines bears the mark •. Then you may append the line "ψ " to the end of $\boldsymbol{D}$.

(55) CONJUNCTION INTRODUCTION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula θ and a line with the formula ψ (in either order). Suppose also that neither of these lines bears the mark •. Then you may append the line "$(\theta \wedge \psi)$ " to the end of $\boldsymbol{D}$.

(56) FIRST CONJUNCTION ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula $(\theta \wedge \psi)$ and that this line is not marked with •. Then you may append the line "θ " to the end of $\boldsymbol{D}$.

(57) SECOND CONJUNCTION ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula $(\theta \wedge \psi)$ and that this line is not marked with •. Then you may append the line "ψ " to the end of $\boldsymbol{D}$.

(58) REITERATION RULE (DERIVED): Suppose that $\boldsymbol{D}$ contains a line with the formula θ and that this line is not marked with •. Then you may append the line "θ " to the end of $\boldsymbol{D}$.

(59) FIRST DISJUNCTION INTRODUCTION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula θ and that this line is not marked with •. Then for any formula ψ whatsoever, you may append the line "$(\theta \vee \psi)$ " to the end of $\boldsymbol{D}$.

(60) SECOND DISJUNCTION INTRODUCTION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula θ and that this line is not marked with $\bullet$. Then for any formula ψ whatsoever, you may append the line "$(\psi \vee \theta)$ " to the end of $\boldsymbol{D}$.

(61) DISJUNCTION ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains lines with the formulas $(\theta \vee \psi)$, $(\theta \rightarrow \chi)$, and $(\psi \rightarrow \chi)$. (The three lines may occur in $\boldsymbol{D}$ in any order.) Suppose also that none of these lines bears the mark $\bullet$. Then you may append the line "χ " to the end of $\boldsymbol{D}$.

(62) NEGATION INTRODUCTION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula $\theta \rightarrow (\psi \wedge \neg\psi)$ not marked by $\bullet$. Then you may append the line "$\neg\theta$ " to the end of $\boldsymbol{D}$.

(63) NEGATION ELIMINATION RULE: Suppose that $\boldsymbol{D}$ contains a line with the formula $\neg\theta \rightarrow (\psi \wedge \neg\psi)$ not marked by $\bullet$. Then you may append the line "θ " to the end of $\boldsymbol{D}$.

(64) DOUBLE NEGATION RULES (DERIVED):

(a) Suppose that $\boldsymbol{D}$ contains a line with the formula θ and that this line is not marked with $\bullet$. Then you may append the line "$\neg\neg\theta$ " to the end of $\boldsymbol{D}$.

(b) Suppose that $\boldsymbol{D}$ contains a line with the formula $\neg\neg\theta$ and that this line is not marked with $\bullet$. Then you may append the line "θ " to the end of $\boldsymbol{D}$.

(65) DEMORGAN (DERIVED): Suppose that $\boldsymbol{D}$ contains a line with the formula $\neg(\theta \vee \psi)$ and that this line is not marked with $\bullet$. Then you may append the line "$\neg\theta \wedge \neg\psi$ " to the end of $\boldsymbol{D}$.

(66) NEGATION INTRODUCTION (DERIVED): Suppose that the last line in $\boldsymbol{D}$ marked by $\circ$ has φ as formula. (Don't use this rule if no line in $\boldsymbol{D}$ is marked by $\circ$.) Suppose also that *either* (a) there are two subsequent

lines in ***D***, neither marked by $\bullet$, and containing the formulas θ and $\neg\theta$, *or* (b) there is a subsequent line in ***D*** unmarked by $\bullet$ containing either $\theta \wedge \neg\theta$ or $\neg\theta \wedge \theta$. Then you may do the following. *First*, from "$\varphi\ \circ$" to the end of ***D***, change the mark of every line to $\bullet$ (if the mark is not $\bullet$ already). *Next*, append the line "$\neg\varphi$ " to the end of ***D***.

(67) NEGATION ELIMINATION (DERIVED): Suppose that the last line in ***D*** marked by $\circ$ has $\neg\varphi$ as formula. (Don't use this rule if no line in ***D*** is marked by $\circ$.) Suppose also that *either* (a) there are two subsequent lines in ***D***, neither marked by $\bullet$, and containing formulas θ and $\neg\theta$, *or* (b) there is a subsequent line in ***D*** unmarked by $\bullet$ containing either $\theta \wedge \neg\theta$ or $\neg\theta \wedge \theta$. Then you may do the following. *First*, from "$\neg\varphi\ \circ$" to the end of ***D***, change the mark of every line to $\bullet$ (if the mark is not $\bullet$ already). *Next*, append the line "φ " to the end of ***D***.

(68) CONTRADICTION RULE (DERIVED): Suppose that ***D*** contains a line with a formula of the form $\varphi \wedge \neg\varphi$, not marked by $\bullet$. Then you may append *any* line marked with the blank.

(69) BICONDITIONAL INTRODUCTION RULE: Suppose that ***D*** contains a line with the formula $(\theta \rightarrow \psi)$ and a line with the formula $(\psi \rightarrow \theta)$ (in either order). Suppose also that neither of these lines bears the mark $\bullet$. Then you may append the line "$\theta \leftrightarrow \psi$ " to the end of ***D***.

(70) FIRST BICONDITIONAL ELIMINATION RULE: Suppose that ***D*** contains a line with the formula θ and a line with the formula $(\theta \leftrightarrow \psi)$ (in either order). Suppose also that neither of these lines bears the mark $\bullet$. Then you may append the line "ψ " to the end of ***D***.

(71) SECOND BICONDITIONAL ELIMINATION RULE: Suppose that ***D*** contains a line with the formula ψ and a line with the formula $(\theta \leftrightarrow \psi)$ (in either order). Suppose also that neither of these lines bears the mark $\bullet$. Then you may append the line "θ " to the end of ***D***.

Chapter 7

Soundness and Completeness

7.1 New notation and chapter overview

Good morning, class. We're very pleased to see you all, with such excitement and impatience written on your faces! Yes, today is the big day. We are going to establish that the derivation rules presented in Chapter 6 do everything we were hoping for. To explain this precisely, we need to introduce some new terminology and review some old. Here's the old, reformulated from Definition (5) in Section 5.1.2.

(1) DEFINITION: If the argument $\Gamma \ / \ \gamma$ is valid — that is, if $[\Gamma] \subseteq [\gamma]$ — then we write $\Gamma \models \gamma$.

For example, you can check that the argument $p \to (q \to r) \ / \ q \to (p \to r)$ is valid. Hence, we can write $p \to (q \to r) \models q \to (p \to r)$. Here is the new notation.

(2) DEFINITION: Let an argument $\Gamma \ / \ \gamma$ be given. If the argument can be derived using the rules introduced in Chapter 6 then we write $\Gamma \vdash \gamma$. If $\Gamma = \emptyset$ then we write this as $\vdash \gamma$.

For example, in Section 6.3.3 we showed how to derive $p \to (q \to r) \ / \ q \to (p \to r)$. Hence, we can write $p \to (q \to r) \vdash q \to (p \to r)$.[1]

If it is false that $\Gamma \models \gamma$, we write $\Gamma \not\models \gamma$, and if it is false that $\models \gamma$ we write $\not\models \gamma$. Similarly, if it is false that $\Gamma \vdash \gamma$, we write $\Gamma \nvdash \gamma$, and if it is false that $\vdash \gamma$ we write $\nvdash \gamma$. By Fact (43) in Section 6.5, $\theta, \psi \in \mathcal{L}$ are interderivable if and only if $\vdash \theta \leftrightarrow \psi$.

In the present chapter we'll show that for all arguments $\Gamma \ / \ \gamma$, $\Gamma \vdash \gamma$ if and only if $\Gamma \models \gamma$. For this purpose, we prove two theorems. First, we'll show that our derivations are *sound* in the sense that every derivable argument is valid. This so-called "soundness theorem" can be formulated as follows.

(3) THEOREM: (Soundness) For all arguments $\Gamma \ / \ \gamma$, if $\Gamma \vdash \gamma$ then $\Gamma \models \gamma$.

[1] The symbol $\vdash$ is named "turnstile" whereas $\models$ is named "double turnstile.

Next we'll show that our derivations are *complete* in the sense that every valid argument is derivable. In other words:

(4) THEOREM: (Completeness) For all arguments Γ / γ, if $\Gamma \models \gamma$ then $\Gamma \vdash \gamma$.

Intuitively, the soundness theorem tells us that our derivation rules tell *only* the truth; they are *trustworthy*. The completeness theorem tell us that the rules tell *all* the truth; they are *informative*. Putting the theorems together yields:

(5) COROLLARY: For all arguments Γ / γ, $\Gamma \vdash \gamma$ if and only if $\Gamma \models \gamma$.

The two theorems are true even when $\Gamma = \emptyset$. In this case, they yield:

(6) COROLLARY: For all $\gamma \in \mathcal{L}$, $\vdash \gamma$ if and only if $\models \gamma$.

We'll attack soundness first, then completeness.[2] Remember we said in Section 6.1 that Chapter 6 was the hardest? We were joking. *This* is the hardest chapter. So you'll be happy to hear that nothing in the rest of the book requires mastery of the proofs of soundness and completeness. You can therefore skip the remainder of the present chapter and pick up the discussion at the start of the next. You'll miss a great ride, though.

7.2 Soundness

7.2.1 Preliminaries

Let a derivation $\boldsymbol{D}$ be given. Let L be any line of $\boldsymbol{D}$. By the "assumption set of L" in $\boldsymbol{D}$ we mean the set of formulas appearing on lines in $\boldsymbol{D}$ that are marked with $\circ$ and occur at or above line L. Consider, for example, the derivations:

[2]Our proof of soundness is tailored to the system we presented in Chapter 5. There doesn't seem to be any straightforward alternative to the approach we'll take. The situation is different for completeness. There are various ways of proceeding. A particularly elegant approach is taken in Mendelson [74, p. 37], based on an argument offered by L. Kalmár in the 1930's.

(7) (a)

1	$p \to q$	$\circ$
2	$q \to r$	$\circ$
3	p	$\circ$
4	q	$\circ$
5	r	
6	s	$\circ$

(b)

1	$p \to q$	$\circ$
2	$q \to r$	$\circ$
3	p	$\circ$
4	q	$\bullet$
5	r	$\bullet$
6	$q \to r$	

The assumption set of line 5 in (7)a is $\{p \to q, q \to r, p, q\}$ whereas the assumption set of line 6 in (7)b) is $\{p \to q, q \to r, p, s\}$. The assumption set of line 2 in both (7)a and (7)b is $\{p \to q, q \to r\}$. (A line with $\circ$ includes its own formula in its assumption set.)

Please recall our use of the term "implies," introduced in Definition (5) of Section 5.1.2. In terms of our new notation, Γ implies γ just in case $\Gamma \models \gamma$.

We'll say that $\boldsymbol{D}$ has "the soundness property" just in case the following is true. For every line ℓ of $\boldsymbol{D}$ not marked by $\bullet$, if φ is the formula on line ℓ then the assumption set of ℓ implies φ. For example, if (7)a has the soundness property then the assumption set of line 5 — namely $\{p \to q, q \to r, p, q\}$ — implies r (which it does).

Now, none of our rules add to the end of derivation a formula marked by $\bullet$ (you can easily check this claim). Therefore, given a derivation of the argument Γ / γ, the derivation will end with γ unmarked by $\bullet$. Now Definition (6) in Section 6.3.1 stipulates that a derivation of Γ / γ leaves the mark $\circ$ only next to members of Γ. It follows that if the derivation has the soundness property then $\Gamma \models \gamma$. To prove Theorem (3) it is therefore enough to show that every derivation has the soundness property. This is precisely what we shall do.

To proceed, we'll first establish that all derivations with one line have the soundness property. Then we will show that extending a derivation using any of our rules maintains the property. This is clearly enough since a derivation $\boldsymbol{D}$ must start with a single line and then grow (if at all) by application of our rules one by one.[3] Notice that we're excluding *derived rules* from the discussion since these were all shown to be just disguised application of the non-derived rules introduced in Section 6.3.

[3]Technically, our proof is by mathematical induction, discussed in Section 2.11.

The case of one line is trivial. For, a single line by itself can only be justified by the Assumption rule (7), of Section 6.3.2, so it has the form:

1	φ	$\circ$	Assumption

The assumption set of 1 is $\{\varphi\}$, which implies φ. So we see that all derivations consisting of just a single line have the soundness property.

Now suppose that $\boldsymbol{D}$ has the soundness property. We'll show that any extension of $\boldsymbol{D}$ using our rules preserves the property. Before getting to specific rules, it will be helpful to make some observations about derivations and validity (they are all easily verified).

(8) FACT: The assumption set of any line in a derivation $\boldsymbol{D}$ includes the assumption set of any line that comes earlier in $\boldsymbol{D}$.

(9) FACT: Let $\Gamma, \Delta \subseteq \mathcal{L}$ and $\gamma \in \mathcal{L}$ be given.[4] Suppose that $\Delta \supseteq \Gamma$. Then if $\Gamma \models \gamma$, also $\Delta \models \gamma$.

(10) FACT: Let $\Gamma \subseteq \mathcal{L}$ and $\gamma, \psi, \chi, \theta \in \mathcal{L}$ be given.[5]

(a) If $\Gamma \models \gamma$ and $\gamma \models \psi$ then $\Gamma \models \psi$.
(b) If $\Gamma \models \gamma$, $\Gamma \models \chi$, and $\{\gamma, \chi\} \models \psi$ then $\Gamma \models \psi$.
(c) If $\Gamma \models \gamma$, $\Gamma \models \chi$, $\Gamma \models \psi$, and $\{\gamma, \chi, \psi\} \models \theta$ then $\Gamma \models \theta$.
(d) If $\Gamma \models \theta \rightarrow \psi$ and $\Gamma \models \theta$ then $\Gamma \models \psi$.
(e) If $\Gamma \models \theta \rightarrow (\psi \wedge \neg\psi)$ then $\Gamma \models \neg\theta$.
(f) If $\Gamma \models \neg\theta \rightarrow (\psi \wedge \neg\psi)$ then $\Gamma \models \theta$.

Now we'll consider the various ways in which $\boldsymbol{D}$ can be extended by our rules. There are 14 possibilities, corresponding to the 14 (non-derived) rules

[4] $\Gamma, \Delta \subseteq \mathcal{L}$ means that both Γ and Δ are subsets of $\mathcal{L}$. In other words, Γ and Δ are sets of formulas. Of course, $\gamma \in \mathcal{L}$ means that γ is some particular formula.

[5] $\gamma, \psi, \chi, \theta \in \mathcal{L}$ means that each of $\gamma, \psi, \chi, \theta$ are members of $\mathcal{L}$. That is, each are formulas.

explained in Section 6.3. We'll go through them one by one (but some can be treated by analogy to others). The only challenging case involves $\rightarrow$ introduction, which we'll treat last. Let $\star$ represent one of the marks $\circ$ or the blank ($\star$ is not $\bullet$).

7.2.2 Assumption

Suppose that $\boldsymbol{D}$ is extended by the Assumption rule (7) of Section 6.3.2. Then the new derivation looks like this:

	$\boldsymbol{D}$	
n	φ	$\circ$

Then the assumption set of line n includes $\{\varphi\}$ so the assumption set of line n implies φ. Hence, the new derivation has the soundness property.

7.2.3 $\rightarrow$ elimination

Suppose that $\boldsymbol{D}$ is extended by the $\rightarrow$ elimination rule (9) of Section 6.3.3. Then the new derivation looks either like this:

various lines ...		
i	$\theta \rightarrow \psi$	$\star$
various lines ...		
j	θ	$\star$
various lines ...		
n	ψ	

or like this:

various lines ...		
i	θ	$\star$
various lines ...		
j	$\theta \rightarrow \psi$	$\star$
various lines ...		
n	ψ	

The two cases are handled virtually identically; we'll just consider the second one. Note that $\boldsymbol{D}$ includes everything before line n. The latter line is the extension of $\boldsymbol{D}$ that we are considering. Let I, J, N be the assumption sets of lines i, j and n respectively. Then $I \cup J \subseteq N$ by Fact (8). Since $\boldsymbol{D}$ has the soundness property, $I \models \theta$ and $J \models \theta \rightarrow \psi$. Therefore, $N \models \theta$ and $N \models \theta \rightarrow \psi$. It follows from Fact(10)d that $N \models \psi$. Thus, $\boldsymbol{D}$ extended by line n has the soundness property.

Until we get to $\rightarrow$ introduction, our treatment of the remaining rules for building derivations will be similar to the case of $\rightarrow$ elimination, just treated. Once the general idea becomes clear, you can skip down to Section 7.2.12, which considers $\rightarrow$ introduction.

7.2.4 $\wedge$ introduction

Suppose that $\boldsymbol{D}$ is extended by the $\wedge$ introduction rule (17) of Section 6.3.4. Then the new derivation looks like this:

various lines ...		
i	φ	$\star$
various lines ...		
j	ψ	$\star$
various lines ...		
n	$\varphi \wedge \psi$	

The original derivation $\boldsymbol{D}$ extends down to the line just before n; the extension is to this latter line. Let I, J, N be the assumption sets of lines i, j, and n, respectively. Then $I \cup J \subseteq N$ by Fact (8). Since $\boldsymbol{D}$ has the soundness property, $I \models \varphi$ and $J \models \psi$. Hence $N \models \varphi$ and $N \models \psi$, from which it follows from Fact (10)b (plus the fact that $\{\varphi, \psi\} \models \varphi \wedge \psi$) that $N \models \varphi \wedge \psi$. Hence, $\boldsymbol{D}$ extended by line n has the soundness property.

7.2.5 $\wedge$ elimination

Suppose that $\boldsymbol{D}$ is extended by the $\wedge$ elimination rule (18) of Section 6.3.4. Then the new derivation looks like this (where $\boldsymbol{D}$ is everything except for the last line n):

various lines ...		
i	$\theta \wedge \psi$	$\star$
various lines ...		
n	θ	

Let I, N be the assumption sets of line i and n, respectively. Then $I \subseteq N$ by Fact (8). Since $\boldsymbol{D}$ has the soundness property, $I \models \theta \wedge \psi$. Hence, $N \models \theta \wedge \psi$, from which it follows from Fact (10)a (plus the fact that $\theta \wedge \psi \models \theta$) that $N \models \theta$. Hence, $\boldsymbol{D}$ extended by line n has the soundness property. The $\wedge$ elimination rule (19) in Section 6.3.4 is analyzed in the same way.

7.2.6 $\vee$ introduction

Suppose that $\boldsymbol{D}$ is extended by the $\vee$ introduction rule (23) in Section 6.3.6. Then the new derivation looks like this (where $\boldsymbol{D}$ is everything except for the last line n):

various lines ...		
i	θ	$\star$
various lines ...		
n	$\theta \vee \psi$	

Let I, N be the assumption sets of lines i and n, respectively. Then $I \subseteq N$ by Fact (8). Since $\boldsymbol{D}$ has the soundness property, $I \models \theta$. Hence, $N \models \theta \vee \psi$, from which it follows from Fact (9) that $N \models \theta \vee \psi$. Hence, $\boldsymbol{D}$ extended by line n has the soundness property. The $\vee$ introduction rule (24) in Section 6.3.6 is analyzed in the same way.

7.2.7 $\vee$ elimination

Suppose that $\boldsymbol{D}$ is extended by the $\vee$ elimination rule (24) in Section 6.3.6. Then the new derivation looks like this (where $\boldsymbol{D}$ is everything except for the last line n):

various lines ...		
i	$\theta \vee \psi$	$\star$
various lines ...		
j	$\theta \rightarrow \chi$	$\star$
various lines ...		
k	$\psi \rightarrow \chi$	$\star$
various lines ...		
n	χ	

or perhaps with lines i, j, and k permuted (the same analysis applies). Let I, J, K, N be the assumption sets of lines i, j, k, and n respectively. Then $I \cup J \cup K \subseteq N$ by Fact (8). Since $\boldsymbol{D}$ has the soundness property, $I \models \theta \vee \psi$, $J \models \theta \rightarrow \chi$, and $K \models \psi \rightarrow \chi$. Hence, $N \models \theta \vee \psi$, $N \models \theta \rightarrow \chi$, and $N \models \psi \rightarrow \chi$. Since $\{\theta \vee \psi, \theta \rightarrow \chi, \psi \rightarrow \chi\} \models \chi$, Fact (10)b implies that $N \models \chi$. Hence, $\boldsymbol{D}$ extended by line n has the soundness property.

7.2.8 $\neg$ introduction

Suppose that $\boldsymbol{D}$ is extended by the $\neg$ introduction rule (27) in Section 6.3.7. Then the new derivation looks like this (where $\boldsymbol{D}$ is everything except for the last line n):

various lines ...		
i	$\theta \rightarrow (\psi \wedge \neg\psi)$	$\star$
various lines ...		
n	$\neg\theta$	

Let I, N be the assumption sets of lines i and n, respectively. Then $I \subseteq N$ by Fact (8). Since $\boldsymbol{D}$ has the soundness property, $I \models \theta \rightarrow (\psi \wedge \neg\psi)$. Hence,

$N \models \theta \rightarrow (\psi \wedge \neg\psi)$, from which it follows from Fact (10)e that $N \models \neg\theta$. Hence, ***D*** extended by line n has the soundness property.

7.2.9 ¬ elimination

Suppose that ***D*** is extended by the ¬ introduction rule (28) in Section 6.3.7. Then the new derivation looks like this (where ***D*** is everything except for the last line n):

	various lines ...	
i	$\neg\theta \rightarrow (\psi \wedge \neg\psi)$	$\star$
	various lines ...	
n	θ	

Let I, N be the assumption sets of lines i and n, respectively. Then $I \subseteq N$ by Fact (8). Since ***D*** has the soundness property, $I \models \neg\theta \rightarrow (\psi \wedge \neg\psi)$. Hence, $N \models \neg\theta \rightarrow (\psi \wedge \neg\psi)$, from which it follows from Fact (10)f that $N \models \theta$. Hence, ***D*** extended by line n has the soundness property.

7.2.10 ↔ introduction

Suppose that ***D*** is extended by the ↔ introduction rule (36) in Section 6.3.8. Then the new derivation looks like this (where ***D*** is everything except for the last line n):

various lines ...		
i	$\theta \rightarrow \psi$	$\star$
various lines ...		
j	$\psi \rightarrow \theta$	$\star$
various lines ...		
n	$\theta \leftrightarrow \psi$	

or else lines i and j are permuted, which changes nothing essential. Let I, J, N be the assumption sets of lines i, j, and n, respectively. Then $I \cup J \subseteq N$ by Fact

(8). Since $\boldsymbol{D}$ has the soundness property, $I \models \theta \rightarrow \psi$ and $J \models \psi \rightarrow \theta$. Hence $N \models \theta \rightarrow \psi$ and $N \models \psi \rightarrow \theta$, from which it follows from Fact (10)b (plus the fact that $\{\theta \rightarrow \psi, \psi \rightarrow \theta\} \models \theta \leftrightarrow \psi$) that $N \models \theta \leftrightarrow \psi$. Hence, $\boldsymbol{D}$ extended by line n has the soundness property.

7.2.11 $\leftrightarrow$ elimination

Suppose that $\boldsymbol{D}$ is extended by the $\leftrightarrow$ elimination rule (37) of in Section 6.3.8. Then the new derivation looks like this (where $\boldsymbol{D}$ is everything except for the last line n):

various lines ...		
i	$\theta \rightarrow \psi$	$\star$
various lines ...		
j	θ	$\star$
various lines ...		
n	ψ	

or else i and j are permuted. Let I, J, N be the assumption sets of line i, j and n, respectively. Then $I \cup J \subseteq N$ by Fact (8). Since $\boldsymbol{D}$ has the soundness property, $I \models \theta$ and $J \models \theta \leftrightarrow \psi$. Hence, $N \models \theta$ and $N \models \theta \leftrightarrow \psi$, from which it follows from Fact (10)b (plus the fact that $\{\theta \leftrightarrow \psi, \theta\} \models \psi$) that $N \models \psi$. Hence, $\boldsymbol{D}$ extended by line n has the soundness property. The other $\leftrightarrow$ elimination rule (38) of Section 6.3.8 is handled in the same way.

7.2.12 $\rightarrow$ introduction

At last, suppose that $\boldsymbol{D}$ is extended by the $\rightarrow$ introduction, rule (8) in Section 6.3.3. Then $\boldsymbol{D}$ can be pictured as

various lines ...		
i	θ	$\circ$
various lines ...		
j	ψ	

where i is the last line marked with $\circ$. An application of $\rightarrow$ introduction extends $\boldsymbol{D}$ to:

various lines ...		
i	θ	$\bullet$
various lines marked with $\bullet$ ...		
j	ψ	$\bullet$
n	$\theta \rightarrow \psi$	

Let I, J be the assumption sets of line i and j in $\boldsymbol{D}$, and let N be the assumption set of line n in the extended derivation. Because $\boldsymbol{D}$ has the soundness property, $J \models \psi$. It follows from Theorem (20) of Section 5.2.2 that $J - \{\theta\} \models \theta \rightarrow \psi$. Now, $N = J - \{\theta\}$ because i is the *last* line marked with $\circ$ in $\boldsymbol{D}$. [If j were marked with $\circ$ then $\rightarrow$ introduction requires i and j to be identical; see Rule (8) in Section 6.3.3.] Hence, $N \models \theta \rightarrow \psi$ so $\boldsymbol{D}$ extended and modified by $\rightarrow$ introduction has the soundness property.

And that's all there is to it! No matter how we construct a derivation, it has the soundness property. Let's repeat how this bears on Theorem (3) (Soundness). If φ is the formula on the last line of a derivation then φ is not marked by $\bullet$, and the assumption set Γ of this line implies φ. By Definition (6) in Section 6.3.1, we derive the argument Γ / φ only when we've produced a derivation ending with φ (unmarked by $\bullet$), and the lines marked by $\circ$ in the derivation are included in Γ. Hence, if $\Gamma \vdash \varphi$ then $\Gamma \models \varphi$. This proves Theorem (3).

That was pretty simple, wasn't it? The proof of completeness is more complicated, and requires some prior theorems of interest in their own right. Before tackling the major results, we record a simple one here.

(11) FACT: Let $\psi, \varphi \in \mathcal{L}$ be interderivable. Then $[\psi] = [\varphi]$.

Proof: Suppose that ψ, φ are interderivable. By the Soundness Theorem (3), $\psi \models \varphi$ and $\varphi \models \psi$. Hence by Definition (5) of Section 5.1.2, $[\psi] = [\varphi]$. ∎

7.3 The Replacement Theorem

Let us recall the concept of "interderivability," defined in Section 6.5. Formulas θ, ψ are interderivable just in case both $\theta \vdash \psi$ and $\psi \vdash \theta$ hold. From Fact (43) in the same section, we know that θ and ψ are interderivable just in case their biconditional is derivable, that is, just in case $\vdash \theta \leftrightarrow \psi$. Of course, a formula is interderivable with itself. That is, $\varphi \leftrightarrow \varphi$ for all $\varphi \in \mathcal{L}$.

Now suppose that $\theta, \psi \in \mathcal{L}$ are interderivable, and consider a formula φ that has θ as subformula. For example:

θ	:	$(p \wedge q)$
ψ	:	$(q \wedge p)$
φ	:	$(p \wedge q) \rightarrow r$

Denote by φ^* the result of replacing all occurrences of θ in φ by ψ. In our example, φ^* is $(q \wedge p) \rightarrow r$. Then we expect φ and φ^* also to be interderivable. This is true in our example; you can easily demonstrate:

$$\vdash ((p \wedge q) \rightarrow r) \;\leftrightarrow\; ((q \wedge p) \rightarrow r)$$

We will now prove that our example represents the general case.

(12) THEOREM: (REPLACEMENT) Let $\varphi, \theta, \psi \in \mathcal{L}$ be given, and suppose that $\vdash \theta \leftrightarrow \psi$. Let φ^* be the result of replacing all occurrences (if any) of θ in φ by ψ. Then $\vdash \varphi \leftrightarrow \varphi^*$.

To prove Theorem (12) we distinguish two cases. The first case is that θ and φ are identical. Then φ^* is ψ. Therefore, φ and θ are interderivable (since every formula of the form $\chi \leftrightarrow \chi$ is derivable, as you know); θ and ψ are interderivable (because we assumed this in the statement of the theorem); and ψ and φ^* are interderivable (because φ^* is just ψ, and to repeat ourselves, every formula is interderivable with itself). By the transitivity of interderivability — which is Corollary (44) in Section 6.5 — it follows that φ and φ^* are interderivable. This is what we're trying to prove.

The second case is that θ and φ are not identical. Since we'll recur to this assumption several times below, let us record it.

(13) θ and φ are not the same formula.

From (13), with the help of Fact (19) in Section 3.6, we infer:

(14) The result of replacing all occurrences (if any) of θ by ψ in φ is the same as the result of replacing all occurrences (if any) of θ by ψ in the principal subformulas of φ.

Recall from Section 3.6 that the principal subformulas of a conjunction are its two conjuncts, the principal subformulas of a disjunction are its two disjuncts, etc. To illustrate (14), let φ be $r \rightarrow \neg(p \vee r)$, let θ be $(p \vee r)$, and let ψ be $(r \vee p)$. Then, θ and φ are not the same formula, and we expect (14) to hold. Indeed, the result of replacing all occurrences of $(p \vee r)$ by $(r \vee p)$ in $r \rightarrow \neg(p \vee r)$ is $r \rightarrow \neg(r \vee p)$, which is the same as replacing all occurrences of $(p \vee r)$ by $(r \vee p)$ in r (there aren't any such occurrences) and replacing all occurrences of $(p \vee r)$ by $(r \vee p)$ in $\neg(p \vee r)$. For another example, let φ be $\neg(p \wedge \neg q)$, let θ be $\neg q$, and let ψ be $\neg\neg\neg q$. Then the result of replacing all occurrences of $\neg q$ by $\neg\neg\neg q$ in $\neg(p \wedge \neg q)$ is the same as the result of replacing all occurrences of $\neg q$ by $\neg\neg\neg q$ in $(p \wedge \neg q)$. We think of the latter replacement as occurring within $\neg(p \wedge \neg q)$ as a whole, which is why replacing $\neg q$ by $\neg\neg\neg q$ in $(p \wedge \neg q)$ yields $\neg(p \wedge \neg\neg\neg q)$. Similarly, the replacements inside the principal subformulas of other kinds of formulas occur *in situ*.

Using (13) and (14), we now proceed to prove Theorem (12) by mathematical induction on the number of connectives that appear in φ. (For mathematical induction, see Section 2.11.) Let φ, θ, ψ and φ^* be as described in the hypothesis of the theorem.[6] We're trying to prove:

(15) $\vdash \varphi \leftrightarrow \varphi^*$.

[6] By this is meant: Let $\varphi, \theta, \psi \in \mathcal{L}$ be given, suppose that $\vdash \theta \leftrightarrow \psi$, and let φ^* be the result of replacing all occurrences (if any) of θ in φ by ψ. In general, the "hypothesis" of a theorem is everything that is assumed, prior to stating the theorem's claim.

Let n be the number of connectives in φ.

Base case: Suppose that $n = 0$, in other words, suppose that φ has no connectives. Then φ is a variable, say, v. We know that θ does not occur in v, since otherwise θ would be v hence θ would be φ, and we ruled this out by (13). Hence (since θ does not occur in v), φ^* is just φ, in other words, φ^* is v. To illustrate, the situation might be as follows.

φ	:	q
θ	:	$r \vee q$
ψ	:	$q \vee r$
φ^*	:	q

So $\vdash \varphi \leftrightarrow \varphi^*$ is true since it amounts to $\vdash v \leftrightarrow v$ for some variable v; and this is just a special case of the fact that every formula is interderivable with itself. This establishes (15) in case n (the number of connectives in φ) is zero.

Now suppose that the number n of connectives in φ is $k + 1$ for some $k \geq 0$. Our work on the base case allows us to assume the following.

(16) INDUCTIVE HYPOTHESIS: Let $\alpha, \theta, \psi \in \mathcal{L}$ be given, and suppose that $\vdash \theta \leftrightarrow \psi$. Suppose also that the number of connectives in α is k or less. Let α^* be the result of replacing all occurrences (if any) of θ in α by ψ. Then $\vdash \alpha \leftrightarrow \alpha^*$.

From (16) we must establish (15). This will prove Theorem (12) since the number of connectives in φ is either zero or greater than zero. Since φ has at least one connective, its principal connective must be one of $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$. (For "principal connective" see Section 3.6.) We consider these possibilities in turn.

φ is a negation. So φ can be represented as $\neg\alpha$, where α has k connectives. [For example, φ might be $\neg(p \wedge \neg q)$. In this case, α is $p \wedge \neg q$.] Let α^* be the result of replacing all occurrences (if any) of θ in α by ψ. [For example, suppose that α is $p \wedge \neg q$, θ is $\neg q$ and ψ is $\neg q \vee \neg q$. Then α^* is $p \wedge (\neg q \vee \neg q)$.] By (16), $\vdash \alpha \leftrightarrow \alpha^*$. [For example, $\vdash (p \wedge \neg q) \leftrightarrow (p \wedge (\neg q \vee \neg q))$.] Now, φ is not θ. [See (13), above.] Hence, by (14), the result of replacing θ by ψ in φ is the result of replacing θ by

ψ in α and then placing $\neg$ in front of α. In other words, φ^* is just $\neg\alpha^*$. [In our example, φ^* is $\neg(p \wedge (\neg q \vee \neg q))$.] Thus, (15) follows directly from the general fact that $\vdash \alpha \leftrightarrow \alpha^*$ implies $\vdash \neg\alpha \leftrightarrow \neg\alpha^*$. [For example, $\vdash (p \wedge \neg q) \leftrightarrow (p \wedge (\neg q \vee \neg q))$ implies $\vdash \neg(p \wedge \neg q) \leftrightarrow \neg(p \wedge (\neg q \vee \neg q))$ This general fact was demonstrated in (45) of Section 6.6.

φ is a conjunction. So φ can be represented as $\alpha_1 \wedge \alpha_2$, where α_1 and α_2 each have k or fewer connectives. [For example, φ might be $((q \vee r) \wedge r) \wedge (q \rightarrow s)$. In this case α_1 is $(q \vee r) \wedge r$ and α_2 is $q \rightarrow s$.] Let α_1^* be the result of replacing all occurrences (if any) of θ in α_1 by ψ. Similarly, let α_2^* be the result of replacing all occurrences (if any) of θ in α_2 by ψ. [For example, suppose θ is $(q \vee r)$ and ψ is $(r \vee q)$. Then α_1^* is $(r \vee q) \wedge r$ and α_2^* is $(q \rightarrow s)$ (that is, just α_2 again).] By (16), $\vdash \alpha_1 \leftrightarrow \alpha_1^*$ and $\vdash \alpha_2 \leftrightarrow \alpha_2^*$. [For example, $\vdash (q \vee r) \wedge r \leftrightarrow (r \vee q) \wedge r$, and $\vdash (q \rightarrow s) \leftrightarrow (q \rightarrow s)$.] Now, φ^* is just $\alpha_1^* \wedge \alpha_2^*$; again, we're relying on (13), above. [In our example, φ^* is $((r \vee q) \wedge r) \wedge (q \rightarrow s)$.] Thus, by (14), (15) follows directly from the general fact that $\vdash \alpha_1 \leftrightarrow \alpha_1^*$ together with $\vdash \alpha_2 \leftrightarrow \alpha_2^*$ imply $\vdash (\alpha_1 \wedge \alpha_2) \leftrightarrow (\alpha_1^* \wedge \alpha_2^*$.) [For example, $\vdash (q \vee r) \wedge r \leftrightarrow (r \vee q) \wedge r$ together with $\vdash (q \rightarrow s) \leftrightarrow (q \rightarrow s)$ implies $\vdash ((q \vee r) \wedge r) \wedge (q \rightarrow s) \leftrightarrow ((r \vee q) \wedge r) \wedge (q \rightarrow s)$.] This general fact was demonstrated in (46) of Section 6.6.

The cases for $\vee$, $\rightarrow$, and $\leftrightarrow$ are similar to the case of $\wedge$. We'll nonetheless plod through the matter for each of the remaining connectives; please forgive the longwindedness. If you're already clear about how the argument goes, just skip down to Section 7.4.

φ is a disjunction. So φ can be represented as $\alpha_1 \vee \alpha_2$, where α_1 and α_2 each have k or fewer connectives. Let α_1^* be the result of replacing all occurrences (if any) of θ in α_1 by ψ. Similarly, let α_2^* be the result of replacing all occurrences (if any) of θ in α_2 by ψ. By (16), $\vdash \alpha_1 \leftrightarrow \alpha_1^*$ and $\vdash \alpha_2 \leftrightarrow \alpha_2^*$. Now, φ^* is just $\alpha_1^* \vee \alpha_2^*$ [we're relying again on (13)]. Thus, using (14), (15) follows directly from the general fact that $\vdash \alpha_1 \leftrightarrow \alpha_1^*$ together with $\vdash \alpha_2 \leftrightarrow \alpha_2^*$ imply $\vdash (\alpha_1 \vee \alpha_2) \leftrightarrow (\alpha_1^* \vee \alpha_2^*$.) This general fact was recorded in (47)a of Section 6.6.

φ is a conditional. So φ can be represented as $\alpha_1 \rightarrow \alpha_2$, where α_1 and α_2 each have k or fewer connectives. Let α_1^* be the result of replacing all occurrences (if any) of θ in α_1 by ψ. Similarly, let α_2^* be the result of replacing all occurrences (if any) of θ in α_2 by ψ. By (16), $\vdash \alpha_1 \leftrightarrow \alpha_1^*$ and $\vdash \alpha_2 \leftrightarrow \alpha_2^*$. Now, φ^* is just $\alpha_1^* \rightarrow \alpha_2^*$

[see (13)]. Thus, relying once more on (14), (15) follows directly from the general fact that $\vdash \alpha_1 \leftrightarrow \alpha_1^*$ together with $\vdash \alpha_2 \leftrightarrow \alpha_2^*$ imply $\vdash (\alpha_1 \rightarrow \alpha_2) \leftrightarrow (\alpha_1^* \rightarrow \alpha_2^*.)$ This general fact was recorded in (47)b of Section 6.6.

φ is a biconditional. So φ can be represented as $\alpha_1 \leftrightarrow \alpha_2$, where α_1 and α_2 each have k or fewer connectives. Let α_1^* be the result of replacing all occurrences (if any) of θ in α_1 by ψ. Similarly, let α_2^* be the result of replacing all occurrences (if any) of θ in α_2 by ψ. By (16), $\vdash \alpha_1 \leftrightarrow \alpha_1^*$ and $\vdash \alpha_2 \leftrightarrow \alpha_2^*$. Now, φ^* is just $\alpha_1^* \leftrightarrow \alpha_2^*$ [see (13)]. Thus, relying a final time on (14), (15) follows directly from the general fact that $\vdash \alpha_1 \leftrightarrow \alpha_1^*$ together with $\vdash \alpha_2 \leftrightarrow \alpha_2^*$ imply $\vdash (\alpha_1 \leftrightarrow \alpha_2) \leftrightarrow (\alpha_1^* \leftrightarrow \alpha_2^*.)$ This general fact was recorded in (47)c of Section 6.6.

And that's the end of the demonstration of Theorem (12).

7.4 DNF formulas again

Getting ready for the proof of completeness [Theorem (4)] requires us to revisit a discussion in Section 5.6. We there defined a formula to be in *normal disjunctive form* just in case it is a disjunction whose disjuncts are simple conjunctions.[7] For example,

$$(\neg q \wedge r) \vee (r \wedge \neg p \wedge q) \vee (p \wedge r)$$

is in normal disjunctive form since it is a disjunction, and each of its disjuncts are conjunctions composed of variables (like p) or their negations (like $\neg p$). As before, we abbreviate "normal disjunctive form" to "DNF." Notice that we don't bother with most of the parentheses in a DNF formula. For example, we write $(r \wedge \neg p \wedge q)$ in place of $(r \wedge (\neg p \wedge q))$ or $((r \wedge \neg p) \wedge q)$. Similarly, the disjunctions are left in "long" form. (See Section 4.4.3 for more discussion of such forms.)

One important point about DNF formulas is that the negation sign $\neg$ applies only to variables, never to bigger formulas as in $\neg(p \vee q)$ (which is not DNF). It may help you recall our previous discussion of DNF by thinking about some special cases. Each of the following formulas is in DNF.

[7]Recall from Definition (57) in Section 5.6 that a simple conjunction is a variable by itself, a negated variable by itself or a conjunction of variables and negated variables.

$p \vee r \vee \neg q$	q	$r \wedge \neg q \wedge \neg p$	$p \wedge \neg p$	$\neg q \vee q$

In Corollary (61) in Section 5.6, we observed that every formula is logically equivalent to a DNF formula. Here is an analogous fact about derivations.

(17) THEOREM: For every formula φ, there is a DNF formula θ such that $\vdash \varphi \leftrightarrow \theta$.

To prove (17) we'll show how to associate with a given formula φ a chain of formulas $\theta_1 \cdots \theta_n$ with the following properties.

(a) $\vdash \varphi \leftrightarrow \theta_1$

(b) for all $i < n$, $\vdash \theta_i \leftrightarrow \theta_{i+1}$.

(c) θ_n is in DNF.

By Fact (44) in Section 6.5 (which states that interderivability is a transitive relation), this is enough to prove the theorem. So let an arbitrary formula φ be given. We construct the chain θ_i in three steps.

Step 1. Recall that for all $\psi, \chi \in \mathcal{L}$, $\psi \rightarrow \chi$ is interderivable with $\neg\psi \vee \chi$, and $\psi \leftrightarrow \chi$ is interderivable with $(\psi \wedge \chi) \vee (\neg\psi \wedge \neg\chi)$. [See (48) in Section 6.6.] Let θ_1 be the result of replacing in φ the leftmost occurrence of a subformula of form $\psi \rightarrow \chi$ or $\psi \leftrightarrow \chi$ with $\neg\psi \vee \chi$ or $(\psi \wedge \chi) \vee (\neg\psi \wedge \neg\chi)$, respectively. (If there are no such occurrences then θ_1 is just φ.) By the Replacement Theorem (12), $\vdash \varphi \leftrightarrow \theta_1$. We repeat this process to obtain θ_2. That is, θ_2 is the result of replacing in θ_1 the leftmost occurrence of a subformula of form $\psi \rightarrow \chi$ or $\psi \leftrightarrow \chi$ with $\neg\psi \vee \chi$ or $(\psi \wedge \chi) \vee (\neg\psi \wedge \neg\chi)$, respectively. So, $\theta_1 \leftrightarrow \theta_2$. Repeat the foregoing process until you reach a formula without any occurrences of $\rightarrow$ or $\leftrightarrow$. Call this formula θ_k. (Of course, θ_k is just φ if φ contains no conditionals or biconditionals as subformulas.) By Theorem (12), $\vdash \varphi \leftrightarrow \theta_1$, and for all $i \leq k$, $\vdash \theta_i \leftrightarrow \theta_{i+1}$. So let us record:

(18) θ_k is interderivable with φ, and θ_k contains no conditionals or biconditionals as subformulas.

Step 2. Transform θ_k into a new formula θ_{k+j} by pushing all negations to the atomic level. In detail, proceed as follows. Find the first subformula of θ_k of form $\neg\neg\psi$, $\neg(\psi \wedge \chi)$ or $\neg(\psi \vee \chi)$. We'll call this subformula the "leftmost offender" in θ_k. If there is no such offender then $j = 0$, and we go to step 3. Assuming that there *is* a leftmost offender, replace it with ψ, $\neg\chi \vee \neg\psi$ or $\neg\chi \wedge \neg\psi$, respectively. By Fact (48) in Section 6.6, θ_k is interderivable with θ_{k+1}. If θ_{k+1} itself has no offender then we are done ($j = 1$). Otherwise, there will be a leftmost offender in θ_{k+1}. This leftmost offender might have already appeared in θ_k, or else it emerged in the transition from θ_k to θ_{k+1}.[For example, θ_k might be $\neg((p \wedge q) \vee r)$, and is thus converted to $\neg(p \wedge q) \wedge \neg r$. The latter formula contains an offender that doesn't appear in $\neg((p \wedge q) \vee r)$.] In either case, let θ_{k+2} be the result of replacing θ_{k+1}'s leftmost offender by the appropriate equivalent formula mentioned above. Keep going like this. You see that eventually the process must exhaust the supply of offenders, leaving each $\neg$ parked next to a variable. You can also see that each of the new θ's generated is interderivable with the preceding one [by Fact (48) in Section 6.6]. Finally, it is also clear that the foregoing procedure inserts no conditionals or biconditionals into any formula. So, in light of (18), we have:

(19) (a) θ_{k+j} is interderivable with φ,

(b) θ_{k+j} contains no conditionals or biconditionals as subformulas, and

(c) all occurrences of $\neg$ in θ_{k+j} appear next to variables.

Let's do an example that illustrates Steps 1 and 2. Starting with

$$\neg(p \rightarrow (q \rightarrow p)),$$

we first eliminate the leftmost conditional to get

$$\neg(\neg p \vee (q \rightarrow p)).$$

Eliminating the second conditional gives

$$\neg(\neg p \vee (\neg q \vee p)).$$

Moving the negation in one step by the DeMorgan law (48)d (Section 6.6) gives

$$\neg\neg p \wedge \neg(\neg q \vee p),$$

and another application of the Demorgan law gives

$$\neg\neg p \wedge \neg\neg q \wedge \neg p,$$

after which two applications of (48)c (Section 6.6) give the final form:

$$p \wedge q \wedge \neg p.$$

Step 3. The present step ensures that disjunctions and conjunctions are properly placed. The only way a formula can fail to be in DNF after steps 1 and 2 is if it has a conjunction governing a disjunction, that is, a subformula of the form $\gamma \wedge (\chi \vee \psi)$ or $(\chi \vee \psi) \wedge \gamma$. In this case the Distribution laws (48)f, g of Section 6.6 tell us that $(\gamma \wedge (\chi \vee \psi))$ and $((\gamma \wedge \chi) \vee (\gamma \wedge \psi)$ are interderivable [or that $(\chi \vee \psi) \wedge \gamma$ and $(\chi \wedge \gamma) \vee (\psi \wedge \gamma)$ are interderivable]. Repeated applications of the Replacement Theorem thus produces a chain of m interderivable formulas beginning with θ_{k+j} from Step 2, and ending with a formula θ_{k+j+m} which is in DNF. Moreover, $\vdash \theta_{k+j+m} \leftrightarrow \theta_{k+j}$.

Step 3 will be clearer in light of an example. We'll convert

$$((p \leftrightarrow q) \rightarrow q)$$

to DNF. All three steps will be used. To start, we eliminate the conditional to reach

$$\neg(p \leftrightarrow q) \vee q.$$

Replacing the biconditional yields

$$\neg((p \wedge q) \vee (\neg p \wedge \neg q)) \vee q.$$

Moving in the negation with Demorgan leads to

$$(\neg(p \wedge q) \wedge \neg(\neg p \wedge \neg q)) \vee q,$$

and moving in the resulting two negations again yields

$$((\neg p \vee \neg q) \wedge (\neg\neg p \vee \neg\neg q)) \vee q.$$

Two double negation eliminations simplify things a little, to

$$((\neg p \vee \neg q) \wedge (p \vee q)) \vee q,$$

but we still need to apply Distribution to reach

$$(\neg p \wedge (p \vee q)) \vee (\neg q \wedge (p \vee q)) \vee q,$$

and then two more Distribution steps to reach our final form

$$(\neg p \wedge p) \vee (\neg p \wedge q) \vee (\neg q \wedge p) \vee (\neg q \wedge q) \vee q.$$

In the foregoing example, we made all uses of Step 1 before any uses of the remaining steps; similarly, we finished up with Step 2 before making any use of Step 3. Such an orderly procedure is always possible because Steps 2 and 3 never introduce $\rightarrow$ or $\leftrightarrow$ into a formula, and Step 3 never displaces a negation. In light of the Replacement Theorem, all the transformations authorized by Steps 1 - 3 involve interderivable formulas. This concludes the proof of Theorem (17).

The only way you're going to understand the proof is to apply its three steps to some formulas, converting them to DNF form.

(20) EXERCISE: For each of the formulas φ below, find a DNF formula θ such that $\vdash \varphi \leftrightarrow \theta$ [thereby illustrating Theorem (17)].

(a) $\neg\neg r \rightarrow (r \vee q)$

(b) $\neg(p \rightarrow q) \rightarrow (q \leftrightarrow p)$

(c) $(p \rightarrow q) \rightarrow r) \rightarrow (q \rightarrow r)$.

7.5 Completeness

At last! We're ready to finish our proof of Theorem (4). To recall what the theorem says, let $\Gamma \ / \ \gamma$ be a valid argument, that is, $\Gamma \models \gamma$. The Completeness Theorem asserts that under these circumstances, $\Gamma \vdash \gamma$, that is, there is a derivation of γ from Γ. We'll now prove this fact, relying on Theorem (17) about DNF form [which, in turn, relies on the Replacement Theorem (12), which in turn relies on the facts proved in Section 6.6]. In particular, we will describe a method that produces a derivation of γ from Γ. This is enough to show that $\Gamma \vdash \gamma$, thereby proving the Completeness Theorem. But we warn you in advance: typically you wouldn't want to use the derivation we'll produce. The relentless method we'll rely on is more thorough than efficient. Often you'll be able to find a shorter derivation on your own.

To prove Theorem (4), suppose that $\varphi_1 \cdots \varphi_n \models \gamma$. (Here, the φ_i's are the members of the set Γ of premises in the argument $\Gamma \ / \ \gamma$ mentioned above.) We must show that $\varphi_1 \cdots \varphi_n \vdash \gamma$. By Theorem (27) in Section 5.2.2, $\varphi_1 \cdots \varphi_n \models \gamma$ implies $\models (\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma$. We will show:

(21) $\models (\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma$ implies $\vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma$.

By (51) of Section 6.6, $\vdash (\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma$ implies $\varphi_1 \cdots \varphi_n \vdash \gamma$, which is what we're trying to demonstrate. So, all that remains is to prove (21). For this purpose, the pivotal step is to demonstrate:

(22) $p \wedge \neg p$ is derivable from $\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma)$.

With (22) in hand, we can construct a derivation for (21) as follows. The derivation starts with the assumption $\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma)$. By (22), the derivation can be extended to $p \wedge \neg p$. By our derived negation elimination rule (32) in Section 6.3.7, we may cancel the assumption, and extend the derivation to $(\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma$. And that's the end. So now all we have to do is demonstrate (22).

So suppose that $\models (\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma$, and picture a derivation that begins with $\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \gamma)$ as assumption. By Theorem (17), there is a DNF formula ψ that is interderivable with $\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma)$. Since $\models (\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma$, its negation $\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma)$ is unsatisfiable; hence $[\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma)] = \emptyset$. Therefore, by Fact (11) in Section 7.2.12, $[\psi]$ = $\emptyset$; that is, ψ is unsatisfiable. Since ψ is in DNF, it follows from Fact (63) in Section 5.6 that every simple conjunction in ψ is a contradictory simple conjunction. Thus, for every simple conjunction χ in ψ there is a variable v such that $\chi \vdash v \wedge \neg v$ (this relies on Conjunction Elimination). Hence, by Fact (49) in Section 6.6, every simple conjunction in ψ derives $p \wedge \neg p$. It then follows immediately from Fact (50) in Section 6.6 that $\psi \vdash p \wedge \neg p$. So, we've constructed a derivation with the assumption $\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma)$ that ends with $p \wedge \neg p$. Negation Elimination then allows us to write $(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma$. The thing can now be pictured this way:

1	$\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma)$	•
	various lines ...	
j	ψ	•
	more lines ...	
k	$p \wedge \neg p$	•
$k+1$	$(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma$	

If the argument to be derived has the form $\{\varphi_1, \cdots, \varphi_n\} \vdash \gamma$, we then use $\to$ elimination to finish with:

1	$\neg((\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma)$	•
	various lines ...	
j	ψ	•
	more lines ...	
k	$p \wedge \neg p$	•
$k+1$	$(\varphi_1 \wedge \cdots \wedge \varphi_n) \to \gamma$	
$k+2$	φ_1	○
	$n-2$ similar lines ...	
$k+n+1$	φ_n	○
	messing around with $\wedge$ introduction ...	
$k+n+1+\ell$	$(\varphi_1 \wedge \cdots \wedge \varphi_n)$	
$k+n+1+\ell+1$	γ	

This establishes (22), and finishes the demonstration of the Completeness Theorem (4).

Are you ready for an example? We will use the method described in the proof of Theorem (4) to demonstrate: $\{\neg p, q \rightarrow (p \wedge r)\} \vdash \neg q$. First, we convert the argument into its "conditional" form, namely: $(\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg q$. The first line of our derivation will therefore be the negation of this conditional:

1	$\neg((\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg q)$ $\circ$	Assumption

The derivation will now be extended down to the DNF formula that is equivalent to line (1). We'll follows steps 1 - 3 in the proof of Theorem (17) above. At the right of each line, we'll indicate what we're doing.

1	$\neg((\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg q)$ $\circ$	Assumption
2	$\neg((\neg p \wedge (\neg q \vee (p \wedge r))) \rightarrow \neg q)$	Replacing first $\rightarrow$
3	$\neg(\neg(\neg p \wedge (\neg q \vee (p \wedge r))) \vee \neg q)$	Replacing next $\rightarrow$
4	$\neg\neg(\neg p \wedge (\neg q \vee (p \wedge r))) \wedge \neg\neg q)$	moving $\neg$ by DeMorgan
5	$(\neg p \wedge (\neg q \vee (p \wedge r))) \wedge q)$	removing $\neg\neg$
6	$((\neg p \wedge \neg q) \vee (\neg p \wedge p \wedge r)) \wedge q$	by distribution
7	$((\neg p \wedge \neg q \wedge q) \vee (\neg p \wedge p \wedge r \wedge q))$	again by distribution

The foregoing array is not an official derivation since it uses the Replacement Theorem (12) as a derived rule. It could be expanded to more elementary steps if desired. At line (7) we've reached the DNF formula that our method squeezes out of the premise. You see that both its disjuncts are contradictory simple conjunctions. From each of these disjuncts we'll now deduce $p \wedge \neg p$. Then we'll use $\vee$-elimination to write $p \wedge \neg p$ free of any assumptions other than line 1. This will allow us to discharge the first line in favor of $(\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg q)$, which is what we wish to derive. Here is a bit more. Lines 1 - 7 are the same as above; some of the marks $\bullet$ start off as $\circ$ (before they are cancelled.)

1	$\neg((\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg q)$	$\circ$	Assumption
2	$\neg((\neg p \wedge (\neg q \vee (p \wedge r))) \rightarrow \neg q)$		Replacing first $\rightarrow$
3	$\neg(\neg(\neg p \wedge (\neg q \vee (p \wedge r))) \vee \neg q)$		Replacing next $\rightarrow$
4	$\neg\neg(\neg p \wedge (\neg q \vee (p \wedge r))) \wedge \neg\neg q)$		moving $\neg$ by DeMorgan
5	$(\neg p \wedge (\neg q \vee (p \wedge r))) \wedge q)$		removing $\neg\neg$
6	$((\neg p \wedge \neg q) \vee (\neg p \wedge p \wedge r)) \wedge q$		by distribution
7	$((\neg p \wedge \neg q \wedge q) \vee (\neg p \wedge p \wedge r \wedge q))$		again by distribution
8	$(\neg p \wedge \neg q \wedge q)$	$\bullet$	Assumption
9	$\neg(p \wedge \neg p)$	$\bullet$	Assumption
10	$\neg q \wedge q$	$\bullet$	$\wedge$ elimination, 8
11	$p \wedge \neg p$	$\bullet$	$\neg$ Elimination (derived), 10
12	$(\neg p \wedge \neg q \wedge q) \rightarrow (p \wedge \neg p)$		$\rightarrow$ Introduction, 8, 14
13	$(\neg p \wedge p \wedge r \wedge q)$	$\bullet$	Assumption
14	$\neg p$	$\bullet$	$\wedge$ Elimination, 13
15	p	$\bullet$	$\wedge$ Elimination, 13
16	$p \wedge \neg p$	$\bullet$	$\wedge$ introduction, 17, 18
17	$(\neg p \wedge p \wedge r \wedge q) \rightarrow (p \wedge \neg p)$		$\rightarrow$ Introduction, 16, 19
18	$p \wedge \neg p$		$\vee$ Elimination, 7, 15, 20

The explicit contradiction at line 18 is what we've been aiming at. Via the derived rule for negation elimination [see (32) in Section 6.3.7], we can remove the negation sign from $\neg((\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg q)$ in line 1, and cancel it as an assumption. At the end, the entire proof looks as follows.

1	$\neg((\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg q)$	•	Assumption
2	$\neg((\neg p \wedge (\neg q \vee (p \wedge r))) \rightarrow \neg q)$	•	Replacing first $\rightarrow$
3	$\neg(\neg(\neg p \wedge (\neg q \vee (p \wedge r))) \vee \neg q)$	•	Replacing next $\rightarrow$
4	$\neg\neg(\neg p \wedge (\neg q \vee (p \wedge r))) \wedge \neg\neg q)$	•	moving $\neg$ by DeMorgan
5	$(\neg p \wedge (\neg q \vee (p \wedge r))) \wedge q)$	•	removing $\neg\neg$
6	$((\neg p \wedge \neg q) \vee (\neg p \wedge p \wedge r)) \wedge q$	•	by distribution
7	$((\neg p \wedge \neg q \wedge q) \vee (\neg p \wedge p \wedge r \wedge q))$	•	again by distribution
8	$(\neg p \wedge \neg q \wedge q)$	•	Assumption
9	$\neg(p \wedge \neg p)$	•	Assumption
10	$\neg q \wedge q$	•	$\wedge$ elimination, 8
11	$p \wedge \neg p$	•	$\neg$ Elimination (derived), 10
12	$(\neg p \wedge \neg q \wedge q) \rightarrow (p \wedge \neg p)$	•	$\rightarrow$ Introduction, 8, 14
13	$(\neg p \wedge p \wedge r \wedge q)$	•	Assumption
14	$\neg p$	•	$\wedge$ Elimination, 13
15	p	•	$\wedge$Elimination, 13
16	$p \wedge \neg p$	•	$\wedge$ introduction, 17, 18
17	$(\neg p \wedge p \wedge r \wedge q) \rightarrow (p \wedge \neg p)$	•	$\rightarrow$ Introduction, 16, 19
18	$p \wedge \neg p$	•	$\vee$ Elimination, 7, 15, 20
19	$(\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg q$		$\neg$ Elimination (derived), 1, 18
20	$\neg p$	○	Assumption
21	$q \rightarrow (p \wedge r)$	○	Assumption
22	$\neg p \wedge (q \rightarrow (p \wedge r))$		$\wedge$ Introduction, 20, 21
23	$\neg q$		$\rightarrow$ Elimination, 19, 22

Our derivation has no more than the two premises of our starting argument as assumptions and it ends with the argument's conclusion.The point of the derivation is to illustrate the general fact that for every valid argument there is a derivation of its conclusion from its premises. Our illustration started with the valid argument $\neg p, q \rightarrow (p \wedge r)$ / $\neg q$.

What happens if we apply our method by mistake to an invalid argument? Is all of the work in trying to find a derivation wasted? To find out, let's try to write a derivation for the *invalid* argument $\neg p, q \rightarrow (p \wedge r)$ / $\neg r$. Applying the procedure outlined in the proof of Theorem (4) yields the following derivation.

1	$\neg((\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg r)$	$\circ$	Assumption
2	$\neg((\neg p \wedge (\neg q \vee (p \wedge r))) \rightarrow \neg r)$		Replacing first $\rightarrow$
3	$\neg(\neg(\neg p \wedge (\neg q \vee (p \wedge r))) \vee \neg r)$		Replacing next $\rightarrow$
4	$\neg\neg(\neg p \wedge (\neg q \vee (p \wedge r))) \wedge \neg\neg r)$		moving $\neg$ by DeMorgan
5	$(\neg p \wedge (\neg q \vee (p \wedge r))) \wedge r)$		removing $\neg\neg$
6	$((\neg p \wedge \neg q) \vee (\neg p \wedge p \wedge r)) \wedge r$		by distribution
7	$((\neg p \wedge \neg q \wedge r) \vee (\neg p \wedge p \wedge r \wedge r))$		again by distribution

As in the previous example, our derivation so far has produced a DNF formula that is equivalent to the formula in line 1. But now we notice that not every disjunct in 7 is a contradictory simple conjunction. The right disjunct is a contradictory simple conjunction but this is not true for the left one. If you go back to Section 5.6 and examine Lemma (63), you will recall that a formula in DNF form is unsatisfiable if and only if all of its disjuncts are contradictory simple conjunctions. Hence, the DNF formula at line 7 *is* satisfiable. Since this formula is interderivable with $\neg((\neg p \wedge (q \rightarrow (p \wedge r)))$, it follows immediately from Fact (11) in Section 7.2.12 that the truth-assignments that satisfy the two formulas are identical. It is clear that any truth assignment that gives truth to r and falsity to p and q satisfies $(\neg p \wedge \neg q \wedge r)$, hence satisfies $((\neg p \wedge \neg q \wedge r) \vee (\neg p \wedge p \wedge r \wedge r))$. So, any such truth-assignment α satisfies $\neg((\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg r)$, hence *fails to satisfy* $(\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg r$, hence makes the left hand side of $(\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg r$ true and the right hand side of $(\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg r$ false. (Right? The only way α can fail to satisfy a conditional $\varphi \rightarrow \psi$ is to satisfy φ and fail to satisfy ψ.) Thus, α is an invalidating truth-assignment for the argument $\neg p, q \rightarrow (p \wedge r)$ / $\neg r$.[8] This invalidating truth-assignment is what our failed derivation of $(\neg p \wedge (q \rightarrow (p \wedge r))) \rightarrow \neg r$ has earned us.

More generally, suppose that we're given an argument A whose derivability we wish to check. We proceed as above by first turning A into a conditional, then constructing a derivation whose first line is the negation of this conditional and whose last line is its DNF equivalent. From the DNF formula it can be seen whether we need to go any further. If all the disjuncts in the DNF formula are contradictory simple conjunctions then we know that the derivation of A can be finished. Otherwise, we pick a disjunct that is not a contradic-

[8] For the concept of an invalidating truth-assignment, see Definition (7) in Section 5.1.2.

tory simple conjunction, and "read off" a truth-assignment that invalidates A. The upshot is that the method used to prove the Completeness Theorem (4) is double edged. Either it produces a derivation of A or an invalidating truth-assignment for A. This sounds great, but as we indicated in Section 6.1, our method is tedious (as the examples suggest). Valid arguments can typically be derived in snappier fashion, and a little thought often reveals an invalidating truth-assignment for an invalid argument without the detour through DNF. The value of the method is thus mainly theoretical, serving to prove the Completeness Theorem (4).

(23) EXERCISE: For each of the arguments below, use the method elaborated above to either write a derivation or produce a counter model.

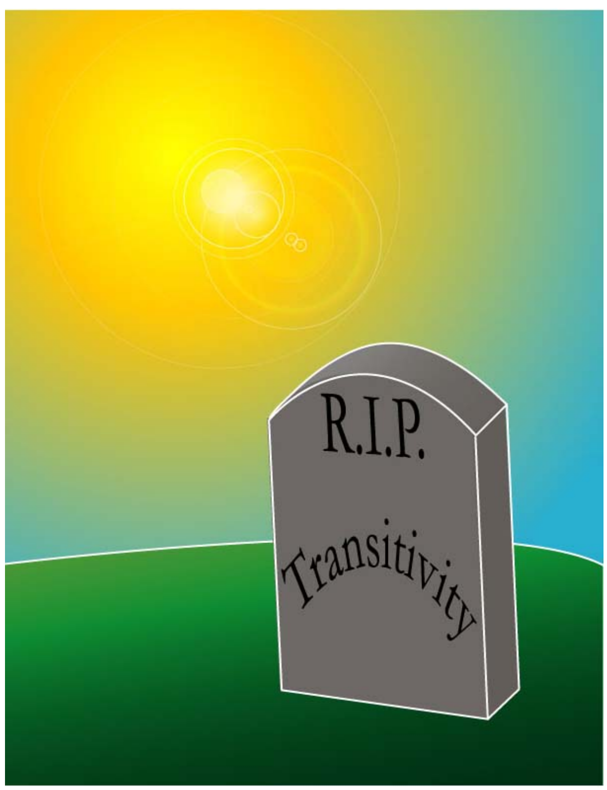
(a) $p \to (q \vee r) \;/\; (p \to r) \vee q$

(b) $p \to (q \to r) \;/\; p \to (q \vee r)$

Aren't you glad to be done with this chapter? That was tough going! Congratulations if you followed it all. But this is no time to rest on your laurels. Fasten your seat belt, we're going places in Chapter 8 . . .

Chapter 8

Problems with conditionals



8.1 Connecting to English

Think about how much we've accomplished. The formal language $\mathcal{L}$ was introduced in Chapter 3, and the meaning of its formulas was explained in Chapter 4. The logical concepts of validity, equivalence, tautology, and contradiction were discussed in Chapter 5. Chapter 6 presented a system of rules for writing derivations, and Chapter 7 showed that the arguments with derivations coincide with the valid arguments. What more could you ask for?

> "Would you kindly explain what all of this has to do with proper reasoning? Proper reasoning, after all, was the topic of the first chapter. How do I use Sentential Logic to decide whether a particular inference is secure?"

Well, if you reason by talking to yourself in the language $\mathcal{L}$ then you are all set! The valid arguments of $\mathcal{L}$ are the ones we've defined. The difficulty, of course, is that few of us actually use $\mathcal{L}$ for ordinary thought, but rather rely on a natural language like English. At least, so it appears to introspection. When thinking, we often sense mental discourse that resembles speech; hardly anyone senses formulas of $\mathcal{L}$. So we are led to the question: What is the relation between $\mathcal{L}$ and a natural language like English? In particular, we would like to know whether there is a simple way to translate sentences of English into formulas of $\mathcal{L}$ so that a translated English argument is secure if and only if its translation into $\mathcal{L}$ is technically valid.

Sorry, there is no such translation schema. Consider the following argument.

(1) There is a song sung by every schoolboy. Therefore, every schoolboy sings some song.

This is a secure inference in English, right? So, we hope to find formulas $\varphi, \psi \in \mathcal{L}$ such that (a) φ "translates" the premise, (b) ψ "translates" the conclusion, and (c) $\varphi \models \psi$. Logicians are agreed that there are no such φ, ψ.[1] Of course,

[1]Bertrand Russell [87] once remarked that whenever all the experts fully endorse some proposition P, he is inclined to believe that P is not *surely* false.

we could arbitrarily choose φ, ψ such that $\varphi \models \psi$. For example, we could map the premise of (1) into $p \wedge q$ and the conclusion into p. But such a choice is not "natural." It provides no clue about how to handle other arguments. More generally, it is not a solution to our problem to assign $p \wedge q$ and q to exactly the secure (one premise) arguments of English, and to assign $p \wedge q$ and r to the non-secure ones. Such a procedure provides no insight into English. It does not help us determine whether an English argument is any good since we must answer that very question prior to choosing formulas for premise and conclusion.

Rather, we were hoping for a translation scheme that seems to preserve the meaning of English sentences, or at least enough of the meaning to make it clear why the inference is secure (or not). Such a scheme would assign to a given sentence of English a formula of $\mathcal{L}$ that represents some or all of its deductive potential. Consider, for example, the following argument.

(2) Either John won the lottery or he inherited his uncle's fortune. John did not win the lottery. Therefore, John inherited his uncle's fortune.

It seems natural to represent the two premises of (2) by $p \vee q$ and $\neg p$, and the conclusion by q. This is because the "or" in the first premise appears to mean (at least roughly) what $\vee$ means in $\mathcal{L}$, just as the "not" appears to mean (roughly) what $\neg$ means.[2] When it is then observed that $\{p \vee q, \neg q\} \models q$, we feel that the logic of $\mathcal{L}$ has illuminated the secure character of the inference in (2). In contrast, no such natural translation is available for Argument (1). In particular, none of the terms "there is," "every," or "some" make contact with distinctive formulas in $\mathcal{L}$. The best we can do is translate the premise of (1) into the nondescript formula p and the conclusion into q; such a bland translation avoids imputing logical structure that is absent from the argument. But such translation provides no logical insight since $p \not\models q$.

Recall that by "Sentential Logic" we mean the logic that governs $\mathcal{L}$. We have just seen that Sentential Logic appears to be adequate to analyze (2) but not (1). Indeed, a stronger logic (based on an artificial language that is more complicated than $\mathcal{L}$) is standardly used to analyze arguments like (1). We shall

[2] See the remarks in Section 4.2.2.

not present that stronger logic here, but rather stick with $\mathcal{L}$.[3] So the question that remains is which arguments of English are successfully analyzed by translating them into $\mathcal{L}$.

Sentential Logic is quite successful in analyzing English arguments whose logical status (secure versus non-secure) depends on the words "and," "or," and "not." Argument (2) provides an illustration. Here is another.

(3) John has red hair and once ran for mayor. Therefore, John once ran for mayor and has red hair.

The argument seems secure, and this fact can be explained by representing it as $p \wedge q \models q \wedge p$. Such a translation is natural but notice that it requires recognizing the elliptical suppression of "John" in the second half of each sentence.[4] Only then can we interpret p as "John has red hair," and q as "John once ran for mayor." Similarly, the non-security of the following argument can be understood via the same interpretation of p and q along with the fact that $p \vee q \not\models q \wedge p$.

(4) Either John has red hair or he once ran for mayor. Therefore, John once ran for mayor and has red hair.

Actually, this case requires some further grammatical judgment before it can be translated into $\mathcal{L}$. We must decide whether "he" in the premise refers to John instead of (say) Rudolph Giuliani. That "he" refers to John seems the most natural assumption since no one but John appears in (4). Once this matter is clarified, translation into $\mathcal{L}$ straightforwardly yields the analysis $p \vee q \not\models q \wedge p$. We often rely on such "massaging" of English sentences to map them onto the variables of a presumed translation into $\mathcal{L}$. Which way the massage should go may depend on incidental facts such as the tendency for "John" to name guys. Thus,

[3] The stronger logic is known as "the predicate calculus," "first order logic," or "quantification theory." There are many fine textbooks that present it, such as [54, 29].

[4] The definition [1] of "ellipsis" is "the omission of a word not necessary for the comprehension of a sentence." In (3), the second occurrences of "John" are omitted without loss of intelligibility.

> John has blue eyes and she is married. Therefore, John is married and has blue eyes.

may well be taken as having the invalid structure $p \wedge q$ / $r \wedge p$.

English comes equipped with various syntactic devices that signal the placement of parentheses in formulas that represent sentences. For example, "Either John or both Mary and Paul ran home" goes over into $p \vee (q \wedge r)$. You can see that the use of "either" and "both" are essential to ensure the translation. Without them, the sentence "John or Mary and Paul ran home" might also be naturally translated as $(p \vee q) \wedge r$. These and other subtleties have been discussed in excellent books like [48, 88] so we won't dwell on them here. It suffices in the present discussion to raise one more issue that complicates the otherwise smooth application of Sentential Logic to English arguments that turn on "and," "or," "not," and related phrases. Consider:

(5) John ran the marathon and died. Therefore, John died and ran the marathon.

Superficially, this argument has the same form as (3) yet we might wish to doubt its security. The use of "and" in (5) appears to code temporal sequencing, as if the argument could be paraphrased as:

(6) John ran the marathon and then died. Therefore, John died and then ran the marathon.

It is clear that the sentences of (6) cannot be represented merely as $p \wedge q$ and $q \wedge p$ since the $\wedge$ has no temporal force in $\mathcal{L}$. There are two potential responses to the different characters of arguments (3) and (5). We can accept that (5) is not secure [unlike (3)], and go on to investigate stronger logics that involve temporal notions like "then."[5] The other response is to affirm the security of (5) despite first appearances, hence to deny that (5) means (6). To defend this response, it may be suggested that "John ran the marathon and died" is typically used to *convey the idea* that John ran and then died but that the sentence doesn't

[5]For a survey of progress in this enterprise, see [33].

mean this. Similarly, you might use the sentence "It's freezing in here" to convey to your brother that he should close the window even though the sentence doesn't *mean* that he should close the window. One says that the suggestion to close the window is conveyed *pragmatically* rather than *semantically*. The semantic/pragmatic opposition has been discussed extensively (see [14, 65, 101] and references cited there). Logicians typically acknowledge that $\wedge$ (for example) does not represent every nuance of "and" but naturally translates "and" in many English arguments nonetheless; the question of whether such nuances are semantic or pragmatic is not addressed directly.[6] Let us adopt the same posture, and close our discussion of "and," "or," and "not."

More profound and vexing questions concern arguments in English that involve the expression *if–then–*, as in:

(7) If the sun is shining then the picnic is in full swing. The birds are singing and the sun is shining. Therefore, the picnic is in full swing.

English sentences constructed around *if–then–* [like the premise of (7)] are often called *conditionals*. The *left hand side* of a conditional is the expression between "if" and "then." The *right hand side* is the expression after "then." Thus, the left hand side and right hand side of the conditional in (7) are "the sun is shining" and "the picnic is in full swing," respectively.

The inference in (7) seems secure, and it invites the translation $\{p \rightarrow q, r \wedge p\} \models q$. Under such translation, the *if–then–* in the first premise is represented in $\mathcal{L}$ by $\rightarrow$. The remainder of the present chapter considers the suitability of this representation. But first we must be careful to specify what kind of conditionals are at issue.

8.2 Two kinds of English conditionals

Compare the following sentences (adapted from Adams [6]).

(8) (a) If J. K. Rowling did not write *Harry Potter* then someone else did.

[6] See [99, p. 80], [100, p. 5], [70, Ch. 5] and [62, p. 64] for a sample of views over the years.

(b) If J. K. Rowling had not written *Harry Potter* then someone else would have.

Both sentences are conditionals but they seem to make fundamentally different claims. In particular, (8)a appears to be undeniably true whereas (8)b will likely strike you as dubious (if not downright false). Also, (8)b strongly suggests that J. K. Rowling was indeed the author of *Harry Potter* whereas no such suggestion emerges from (8)a. Finally, the left hand side of (8)a can stand alone as the first sentence in a conversation. That is, you can walk up to someone and blurt out "J. K. Rowling did not write *Harry Potter*." We don't suggest that you actually do this; our point is just that such a sentence makes sense standing alone. In contrast, there is something (even) odd(er) about blurting out "J. K. Rowling had not written *Harry Potter*," which is the left hand side of (8)b. Such an utterance makes it appear that you're engaged in dialogue with an invisible interlocutor, raising the need for medical assistance. Here's another example, without the negations in the left hand sides.

(9) (a) If Jason Kidd was trained in astrophysics then he is the scientist with the best 3-point shot.

(b) If Jason Kidd were trained in astrophysics then he would be the scientist with the best 3-point shot.

The same three distinctions appear to separate the two sentences of (9). Whereas (9)a seems undeniable, (9)b can be disputed (spending your time looking through telescopes might well ruin your shot). Also, (9)b but not (9)a suggests that Kidd failed to receive training in astrophysics. Finally, the left hand side of (9)a can be sensibly asserted in isolation whereas (9)b cannot. In fact, "Jason Kidd were trained in astrophysics" isn't even English.[7]

[7] We've just claimed that indicative but not subjunctive conditionals have left hand sides that can be sensibly asserted in isolation. Michael McDermott has pointed out to us, however, that the distinction may not be so sharp. The indicative conditional "If it rains tomorrow then the grass will grow" has left hand side "it rains tomorrow," which may not be independently assertible. Our view is that "it rains tomorrow" is admittedly marginal but not much worse than "a meteor strikes Earth in 2020" which we judge (with some queasiness) to be OK. Both seem qualitatively better than "J. K. Rowling had not written *Harry Potter*," and "Jason Kidd were trained in astrophysics." You'll have to make up your own mind about these cases.

Let us pause to note that not everyone agrees that sentences like (9)b indicate the falsity of their left hand sides. Consider the following example (adapted from Anderson [7]).

(10) If the victim had taken arsenic then he would have shown just the symptoms that he in fact shows.

The truth of this sentence is often said to strengthen the claim that the victim had taken arsenic rather than weaken it. The present authors, in contrast, find the sentence strange, on the model of the more frankly puzzling example:

(11) If Barbara Bush had voted for her son in 2000 then George W. Bush would have carried Texas.

George W. *did* carry Texas in 2000, just as the victim in (10) *did* show just the symptoms he in fact showed. Yet (11) seems to suggest that Barbara didn't vote for her son, just as (we think) the first suggests that the victim didn't take arsenic after all. We hope you agree with us; if not, you'll have to keep this caveat in mind for the sequel.

It is interesting to think about the grammatical differences between the two sentences in each pair, (8), (9). The distinction between (8)a and (8)b turns on the use of "did not" in the left hand side of the first and "had not" in the left hand side of the second. In the right hand sides, this distinction plays out in the contrast between "did" and "would have." In (9), the left hand sides oppose "was" against "were" (and "is" against "would be"). It is said that (8)a and (9)a exhibit the *indicative* mood whereas (8)b and (9)b exhibit the *subjunctive* mood. In English, the difference in mood is marked by the use of auxiliary verbs that also serve other purposes (e.g., "were" is also the past tense form used with "you"). In many other languages (e.g., Italian) the subjunctive mood is honored with a distinctive form of the verb.[8] Conditionals involving the indicative mood are called "indicative conditionals;" those involving the subjunctive mood are

[8]Thus, in Italian, (8)a can be rendered by:

Se J. K. Rowling non ha scritto *Harry Potter* allora qualcun' altro l'ha scritto.

In contrast (8)b is best translated:

called “subjunctive conditionals”. Some people qualify subjunctive conditionals as “counterfactual,” but we’ll avoid this terminology (preferring syntactic to semantic criteria).

The difference between indicative and subjunctive conditionals shows up in the secure inferences they support. Consider the following contrast (drawn from Adams [5]).

(12) (a) If Jones was present at the meeting then he voted for the proposal.

(b) If Jones had been present at the meeting then he would have voted for the proposal.

Only (12)a seems to follow from:

(13) Everyone present at the meeting voted for the proposal.

The subjunctive conditional (12)b cannot be securely inferred from (13) since the latter sentence says nothing about non-attendees like Jones.

In what follows we shall concern ourselves exclusively with indicative conditionals, not subjunctive. The reason for the choice is that there is little hope of representing subjunctive conditionals successfully in $\mathcal{L}$, our language of Sentential Logic. The conditional $\rightarrow$ of $\mathcal{L}$ is plainly unsuited to this purpose. For one thing, the left hand side of subjunctive conditionals like (8)b and (9)b may not have truth-values in the ordinary sense since (as we saw) they seem not to be sensibly assertible in isolation; in the absence of such truth values, the semantics of $\rightarrow$ [namely, its truth table (18), described in Section 4.2.4] cannot even be applied to subjunctive conditionals. And if we *do* take the left hand sides of subjunctive conditionals to have truth-values then $\rightarrow$ surely gives the wrong interpretation. Consider the following contrast.

(14) (a) If Bill Clinton had touched the ceiling of the Senate rotunda then it would have turned to solid gold.

Se J. K. Rowling non avesse scritto *Harry Potter* allora qualcun’ altro l’avrebbe scritto.

The specialized form “avesse” marks the subjunctive in Italian.

(b) If Bill Clinton had touched the ceiling of the Senate rotunda then it would have remained plaster.

Clearly, (14)a is false and (14)b is true. Yet, if the left hand side of (14)a has a truth-value, it would seem that the value must be *false* since Bill never did touch the ceiling of the Senate rotunda. (We know this.) Now recall (from Section 4.2.2) that according to Sentential Logic, every conditional in $\mathcal{L}$ with false left hand side is true. Thus, Sentential Logic cannot distinguish the truth values of the two sentences in (14) if we try to represent them using $\rightarrow$.[9] In fact, adequately representing subjunctive conditionals requires that the syntax and semantics of $\mathcal{L}$ be considerably enriched, and there are competing ideas about how best to proceed. For an introduction to the issues, see [78, 12, 69].[10]

Perhaps a similar example also discourages us from adopting $\rightarrow$ as a translation of the indicative conditional. Consider this contrast:

(15) (a) If Bill Clinton touched the ceiling of the Senate rotunda then it turned to solid gold.

(b) If Bill Clinton touched the ceiling of the Senate rotunda then it remained plaster.

Since the common left hand side of these conditionals is false, both (15)a,b come out true if we represent them using $\rightarrow$. Our (admittedly faint) intuition is that declaring (15)a,b to be true is more plausible than such a declaration about (14)a,b. But let's agree to leave this issue in abeyance for now (we'll return to it in Chapter 10). The important thing for now is to circumscribe our investigation. In the present work, we stay focussed on indicative conditionals.

[9] Yet other subjunctive conditionals seem to have no truth-value at all, such as:

If Houston and Minneapolis were in the same state then Houston would be a lot cooler.

(We considered this example in Section 1.4). Is the sentence true, or would Minneapolis be a lot warmer? Or would there be a new, very large state? (For more discussion of this kind of case, see [81].)

[10] And see [45] for an anthology of influential articles on the logic of English conditionals.

But we haven't really defined the class of English indicative conditionals. It is tempting to identify them as the sentences with *if–then–* structure that involve the indicative mood. This definition, however, is at once too narrow and too broad. It is too narrow because there are many English sentences that don't involve *if–then–* yet seem to express the same meaning. You've probably noticed that the word "then" can often be suppressed without changing the meaning of an indicative conditional. For example, (16)b seems to express the same thing as (16)a

(16) (a) If humans visit Mars by 2050 then colonies will appear there by 2100.

(b) If humans visit Mars by 2050, colonies will appear there by 2100.

It may not have occurred to you that *both* "if" and "then" are dispensable in conditionals. Consider:

(17) You keep talkin' that way and you're gonna be sorry!

Despite the "and," (17) seems to mean no more nor less than:

(18) If you keep talkin' that way then you're gonna be sorry!

Since (18) is an indicative conditional, perhaps we ought to count (17) as one too. Other conditional-like constructions that don't involve *if–then–* are:

The plane will be late *in the event* (or *in case*) of fog.

The plane will be late *should there happen to be* fog.

The plane will be late *assuming there to be* fog.[11]

Likewise, there are sentences involving *if–then–* and the indicative mood that seem quite different from the indicative conditionals (8)a and (9)a discussed above. Consider, for example:

[11]The grammatical relations among these different constructions are considered in Lycan [69], and references cited there.

(19) (a) If a star is red then it is cooler than average.

(b) If male elks have horns then they are aggressive.

Despite the indicative mood, the left hand sides and right hand sides of these two sentences don't seem to carry truth-values in the usual sense. The left hand side of (19)a does not assert that some particular star is red, nor does the left hand side of (19)b asserts that all male elks have horns. Rather, (19)a seems to assert something equivalent to "every red star is cooler than average," and (19)b seems to assert something like "every male elk with horns is aggressive." These interpretations are suggested by the use of the pronouns "it" and "they" in (19). Yet we don't mean to imply that every use of pronouns in the right hand side excludes the sentence from the class of indicative conditionals. Thus, the sentence "If John studies all night then he'll pass the test" is clearly an indicative conditional since it is paraphrased by the pronoun-free sentence "If John studies all night then John will pass the test." In contrast, it is hard to see how to rid (19)a,b of their pronouns without a change in meaning.

Other uses of *if–then–* yield sentences whose status as indicative conditionals is unclear. Consider:

> If you really want to know, *I'm* the one who added chocolate chips to the baked salmon.

Perhaps this is a genuine conditional in view of its *if–then–* form. Or perhaps it's just masquerading as a conditional inasmuch as its left hand side seems intended merely to communicate attitude (". . . and even if you *don't* want to know, *I'm* the one who did it!").[12]

Let us also note that the form of auxiliaries marking subjunctive conditionals is subject to dialectical variation in America. For example, many people can use the following sentence to mean what we expressed in (9)b.

> If Jason Kidd was trained in astrophysics then he'd be the scientist

[12]Compare: "Let me tell you something, Bud. *I'm* the one who added chocolate chips to the baked salmon." For discussion of a range of such cases, see [69, Appendix].

with the best 3-point shot.[13]

It would be tiresome to track down and classify all the syntactic peculiarities that include or exclude sentences from the class of indicative conditionals that we have in mind. We'll just let (8)a and (9)a serve as paradigm cases, and also note that the left hand side and right hand side of indicative conditionals must be able to stand alone as truth-bearing declarative sentences. That is, both fragments must be either true or false, whether or not the speaker, listener, or reader happens to know which truth-value is the right one.

8.3 Hopes and aspirations for Sentential Logic

With the foregoing qualifications in mind, let us now try to be clear about what we expect from Sentential Logic. We'll do this by formulating a *criterion of adequacy* for logic to serve as a guide to secure inference, or rather, a *partial* guide since we've seen that there are inferences beyond the purview of Sentential Logic.

(20) CRITERION OF ADEQUACY FOR LOGIC: For every argument $\varphi_1 \ldots \varphi_n$ / ψ of $\mathcal{L}$, $\varphi_1 \ldots \varphi_n \models \psi$ if and only if every argument $P_1 \ldots P_n$ / C of English *that is naturally translated into* $\varphi_1 \ldots \varphi_n$ / ψ is secure.

To illustrate, the argument $p \wedge q$ / $q \wedge p$ appears to conform to (20) since (a) $p \wedge q \models q \wedge p$ and (b) every English argument that is naturally translated into

[13] Yankees manager Joe Torre commented on his player Hideki Matsui as follows (quoted in the *New York Times*, 9/22/03).

> If he was anything less than what he is, we aren't near where we are. He's given us such a lift.

In the King's English, Torre's comment comes out to be:

> If Hideki Matsui had skills inferior to those he actually possesses then the Yankees would not be as far ahead in the pennant race as they in fact are. Quite a lift he's given us!

But of course, Kings don't know nuttin' about baseball.

$p \wedge q \;/\; q \wedge p$ is secure. At least, all such English arguments *seem* to be secure; for example, (3) is one such argument. Likewise, $p \vee q \;/\; q \wedge p$ conforms to (20) since (a) it is invalid, and (b) not every English argument that is naturally translated into $p \vee q \;/\; q \wedge p$ is secure; a counterexample is (4).

Notice how slippery Criterion (20) is. If $\varphi_1 \ldots \varphi_n \;/\; \psi$ is valid, we must be content with a just a sample of English counterparts in order build confidence that *all* arguments translatable into $\varphi_1 \ldots \varphi_n \;/\; \psi$ are secure. We have so little handle on English that it's not feasible to *prove* that there are no exceptions. On the other hand, if $\varphi_1 \ldots \varphi_n \;/\; \psi$ is invalid then we are a little better off since just a single non-secure argument of the right form suffices to nail down conformity with (20).

Criterion (20) is slippery also because we haven't been precise about which translations into $\mathcal{L}$ are "natural." This opens a loophole whenever we find a non-secure argument that translates into a validity. We can always complain afterwards that the translation is not natural. Such complaints might be hard to dismiss. We saw above, for example, that whether a sentence is an indicative conditional is often a subtle affair. And the affair is consequential since only indicative conditionals are considered to be "naturally translated" by formulas with $\rightarrow$ as principal connective. No natural translation into $\mathcal{L}$ is recognized for subjunctive conditionals. Even simple cases like (5) (John's death after the marathon) raise knotty questions about natural translation.

Despite the slip and slop in Criterion (20), we shall see that it imposes tough standards on Sentential Logic. There is enough agreement about natural translation to allow different people to be convinced by the same examples much of the time. We hope to convince you of this fact in what follows.

So, at last, we are ready to address the central issue in this chapter. Are indicative conditionals successfully represented by the $\rightarrow$ of Sentential Logic? This is such a nice question that we'll provide two different answers. First, conclusive proof will be offered that Yes, $\rightarrow$ is an appropriate representation of indicative *if–then–*. Next, conclusive proof will be offered that No, $\rightarrow$ is not an appropriate representation of indicative *if–then–*. (Isn't logic great?) Afterwards, we'll try to make sense of this apparent contradiction.

8.4 Indicative conditionals can be represented by $\rightarrow$

8.4.1 Some principles of secure inference

To make our case that *if–then–* can be represented by $\rightarrow$, some more notation will be helpful. Let $A_1 \cdots A_n / \ C$ be an argument in English with premises $A_1 \cdots A_n$ and conclusion C. For example, the argument might be:

(21) A_1**:** If the Yankees lost last night's game then the general manager will be fired within a week.

A_2**:** The Yankees lost last night's game.

Therefore:

C**:** The general manager will be fired within a week.

We write $\{A_1 \cdots A_n\} \Rightarrow C$ just in case it is not possible for all of $A_1 \cdots A_n$ to be true yet C be false. For example, it is impossible for the premises of (21) both to be true without the conclusion being true as well. So for this argument, we write $\{A_1, A_2\} \Rightarrow C$. To reduce clutter, we sometimes drop the braces, writing (for example): $A_1, A_2 \Rightarrow C$.

Our definition of the $\Rightarrow$ relation just symbolizes what we already discussed in Section 1.3 when we outlined the goals of deductive logic. If $\{A_1 \cdots A_n\} \Rightarrow C$ holds, then the inference from $A_1 \cdots A_n$ to C is secure; the truth of the premises *guarantees* the truth of the conclusion. You can see $\Rightarrow$ is the counterpart to $\models$ in Sentential Logic. But the former holds between premises and conclusions written in English whereas the latter holds between premises and conclusions written in $\mathcal{L}$.

Let it be noted that our definition of secure inference sits on a volcano of complex issues. We haven't been clear about the type of "impossibility" or "guarantee" involved in supposedly secure inference. Take the argument with premise

Charles Bronson was a riveting actor

and conclusion

Charles Buchinsky was a riveting actor.

Is this inference secure? Well, it turns out to be impossible for the premise to be true and the conclusion false since Charles Bronson *is* Charles Buchinsky (he wisely changed his name). Yet the status of the inference remains ambiguous (it is guaranteed in one sense but not another). Many other ambiguities could be cited. Rather than enter the inferno of discussion about possibility, we will attempt to rely on a loose and intuitive sense of secure inference. An inference is secure if (somehow) the *meaning* of the premises and conclusion ensures that the former can't be true and the latter false.

Using our new notation, let us formulate some principles that appear to govern secure inference in English. Below, by "sentence" we mean "declarative sentence of English with a determinate truth-value," in accord with our usual convention.

(22) TRANSITIVITY: Let three sentences A, B, C be such that $A \Rightarrow B$ and $B \Rightarrow C$. Then $A \Rightarrow C$.

Right? If it's impossible for A to be true without B being true, and likewise it is impossible for B to be true without C being true, then it is impossible for A to be true without C being true. This seems self-evident to us, but we don't wish to dogmatically impose it on you. If you think we're wrong then you should be cautious about whatever depends on (22) in what follows.

The remaining principles refer to grammatical constructions that mirror some of the syntax of $\mathcal{L}$. Thus, we'll write "*not-*A" to refer to the negation of the English sentence A. To illustrate, if A is "Lions bark" then *not-*A is "Lions don't bark." The syntactic difference between A and *not-*A depends on the particular structure of A. For example, If A were "Lions don't dance," then *not-*A might be "It's not true that Lions don't dance," or perhaps "Lions dance." It suffices for our purposes to allow *not-*A to be any such negation of A. Likewise, A*-and-*B is the result of conjoining sentences A and B with the word "and," or combining them in some equivalent way. Thus, if A is "Lions bark" and B is "Tigers bark," then A*-and-*B is "Lions bark and tigers bark," or perhaps "Lions and tigers bark." The same remarks apply to the notation A*-or-*B. Finally, "*if-*A*-then-*B"

is the *if–then–* sentence with A as left hand side and B as right hand side, or something equivalent. With A and B as before, *if-A-then-B* might be "If lions bark then tigers bark" or "Tigers bark if lions do." Now we make some claims about the foregoing constructions, by announcing some more (putative) principles of English.

(23) DEDUCTION PRINCIPLE FOR ENGLISH: Let three sentences A, B, C be such that $\{A, B\} \Rightarrow C$. Then $A \Rightarrow$ *if-B-then-C*.

To illustrate, let A, B, C be as follows.

A**:** Either Sally will cut out the racket or Sam is going to leave.

B**:** Sally will not cut out the racket.

Therefore:

C**:** Sam is going to leave.

This is a case in which $\{A, B\} \Rightarrow C$ (right?). It illustrates (23) inasmuch as the argument from the premise

> Either Sally will cut out the racket or Sam is going to leave.

to the conclusion

> If Sally will not cut out the racket then Sam is going to leave.

seems secure. In other words: $A \Rightarrow$*if-B-then-C*. The example does not *prove* (23); it only *illustrates* the principle. We don't know how to rigorously prove the principles formulated in this section since they concern English, which no one knows how to formalize. You may nonetheless be persuaded (as seems plausible) that (23) holds in full generality. By the way, (23) resembles Fact (20) in Section 5.2.2 which is often called the "Deduction Theorem" for Sentential Logic.

Moving along, here are some other principles.

(24) CONTRADICTION PRINCIPLE FOR ENGLISH: For every pair A, B of sentences, $\{A, \textit{not-}A\} \Rightarrow B$.

The foregoing principle has already been discussed and justified in Section 5.3.2.

(25) FIRST CONDITIONAL PRINCIPLE FOR ENGLISH: For every pair A, B of sentences, *if-*A*-then-*B $\Rightarrow$ *not-*$(A$*-and-not-*$B)$

We haven't formally introduced the expression *not-*$(A$*-and-not-*$B)$ but it should be transparent by this point. It is the result of negating the English sentence that comes from conjoining via "and" the sentence A and the negation of sentence B. To illustrate, let A be "Lions bark" and B be "Zoo-keepers are amazed." Then (25) asserts that the truth of

(26) If lions bark then zoo-keepers are amazed.

guarantees the truth of

It's not true that lions bark and zoo-keepers are not amazed.

The guarantee stems from the impossibility that both (26) and

Lions bark and zoo-keepers are not amazed.

are true.

(27) SECOND CONDITIONAL PRINCIPLE FOR ENGLISH: For every pair A, B of sentences, A*-or-*B $\Rightarrow$ *if-not-*A*-then-*B.

Thus, if it is true that either whales dance or turtles sing, then it may be securely inferred that if whales don't dance then turtles sing.

(28) DEMORGAN PRINCIPLE FOR ENGLISH: For every pair A, B of sentences, *not-*$(A$*-and-*$B)$ $\Rightarrow$ *not-*A*-or-not-*B.

If A and B are "Lions bark" and "Dogs bark," then (28) asserts — quite plausibly — that the truth of

> It's not true that both lions and dogs bark.

guarantees the truth of

> Either lions don't bark or dogs don't bark.

Principle (28) is the English counterpart of a law of Sentential Logic usually named after Augustus DeMorgan. It was presented in Section 6.3.7.

Finally, we formulate a double-negation principle.

(29) DOUBLE NEGATION PRINCIPLE FOR ENGLISH: Suppose that sentence B contains a sentence of the form *not-not-*A inside of it. Let C be the sentence that results from substituting A for *not-not-*A in C. Then $B \Rightarrow C$.

For example, (29) asserts that the truth of

> Dogs bark and it is not true that sparrows don't fly.

guarantees the truth of

> Dogs bark and sparrows fly.

In this example, A is "Sparrows fly." As a special case (in which B contains nothing else than *not-not-*A), Principle (29) asserts that *not-not-*$A \Rightarrow A$, e.g., that

> It is not true that sparrows don't fly.

allows the secure inference of

> Sparrows fly.

The principles discussed above should all strike you as plausible claims about English, but in fact there is a complication. The double negation principle (29), for example, is open to the following (dumb) counterexample.

(30) A**:** Sparrows fly.

B**:** John said: "It is not true that sparrows don't fly."

C**:** John said: "Sparrows fly."

Even though *not-not-*A occurs inside of B, it is possible for B to be true and C false (John might never express himself concisely). Of course, sentences with internal quotation are not what we had in mind! We were thinking of simple English declarative sentences, the flat-footed kind, reporting straightforward facts (or non-facts). Unfortunately, we don't know how to rigorously define this set of sentences, even though we suspect you understand what set we have in mind. So let us proceed as follows. Consider the set S of declarative English sentences (with determinate truth-values) that *do* satisfy the principles formulated in this section. We hope you agree that S is richly populated and worthy of study. The question animating the present chapter then becomes: Are indicative conditionals *with left hand side and right hand side belonging to* S successfully represented by the $\rightarrow$ of Sentential Logic? Relying on our principles, we'll now present two arguments in favor of an affirmative answer.

8.4.2 First argument showing that indicative conditionals are faithfully represented by $\rightarrow$

Here is the truth-table for conditionals within Sentential Logic, repeated from Section 4.2.4.

(31) TABLE FOR CONDITIONALS:

χ	$\rightarrow$	ψ
T	T	T
T	F	F
F	T	T
F	T	F

Suppose that we were persuaded of the following facts about a given indicative conditional *if-E-then-F*.

(32) (a) If F is true then *if-E-then-F* is true.

(b) If E is false then *if-E-then-F* is true.

(c) If E is true and F is false then *if-E-then-F* is false.

Then we will have shown that *if-E-then-F* is true in exactly the same circumstances in which $E \rightarrow F$ is true, and false in the same circumstances that $E \rightarrow F$ is false. You can see this by examining each line of the truth table (31). The first line reveals that $E \rightarrow F$ is true if both E and F are true. But (32)a asserts that *if-E-then-F* is likewise true in these circumstances (since F is true). The second line of (31) shows that $E \rightarrow F$ is false if E is true and F is false; and this circumstance makes *if-E-then-F* false according to (32)c. The third line of (31) exhibits $E \rightarrow F$ as true if E is false and F is true. But since F is true in this case, (32)a can be invoked once again to show that *if-E-then-F* is true in the same circumstances [we could also have relied on (32)b in this case]. Finally, the fourth line of (31) reveals $E \rightarrow F$ to be true if both E and F are false. In these circumstances E is false, and (32)b states that *if-E-then-F* is true. So it appears to be sufficient to argue in favor of (32) in order to establish:

(33) An indicative conditional *if-E-then-F* is true if and only if $E \rightarrow F$ is true.

Since the security of arguments in English concerns no more than guaranteeing the truth of the conclusion given the truth of the premises, (33) seems to be all we need to justify representing *if–then–* of English by $\rightarrow$ of $\mathcal{L}$.

It remains to convince ourselves of (32), which will convince us of (33), which will convince us that $\rightarrow$ successfully represents *if–then–* in $\mathcal{L}$. But let us first address an issue that might be troubling you.

> "The three claims in (32) are formulated using English *if–then–*. Yet we are in the middle of presenting contradictory claims about the

meaning of this locution. Apparently, we're not yet certain what *if–then–* means, so how can we sensibly discuss the meaning of *if–then–* while using that very meaning in our discussion?"

Several responses can be offered to this excellent question. One is to observe that we could write (32) equally well as:

(34) (a) F is true $\rightarrow$ *if-E-then-F* is true.

(b) E is false $\rightarrow$ *if-E-then-F* is true.

(c) E is true and F is false $\rightarrow$ *if-E-then-F* is false.

The $\rightarrow$ is here interpreted exactly as in $\mathcal{L}$, namely as yielding a true sentence unless the left hand side is true and the right hand side is false. To establish (33) it suffices to establish (34); this can be seen via the same reasoning used above concerning (32). The *if and only if* seen in (33) can likewise be understood as $\leftrightarrow$ in $\mathcal{L}$. Such an interpretation of (33) is enough to underwrite the claim that *if–then–* is suitably represented by $\rightarrow$.

Indeed, *if–then–* in (32) can be understood in several ways without altering its support for (33). For example, we could have written (32) as:

(35) (a) F can't be true without *if-E-then-F* being true.

(b) E can't be false without *if-E-then-F* being true.

(c) E can't be true and F false without *if-E-then-F* being false.

We could still infer (33).

There is another response to your worry about (32) that is worth recording. We are presently trying to discover something about the meaning of *if–then–*, but it has not been doubted that *if–then–* has a definite meaning that is understood (albeit implicitly) by speakers of English. What is wrong with relying on our shared understanding of *if–then–* while discussing it? Similarly, we would not hesitate to use the word "tiger" in discussions of the biological nature of tigers. It would seem odd to question such use of "tiger" on the grounds that we had not yet finished our inquiry. Let us frankly admit to not being sure how far this analogy between "tiger" and *if–then–* can be pushed. But we'll nonetheless continue to freely use *if–then–* locutions in our discussion of *if–then–*. Naturally, we will endeavor to use *if–then–* in a manner consonant with common understanding.

Now, what reason is there to believe the claims of (32)? They all follow from the principles reviewed in Section 8.4.1 above! Consider first (32)a. We argue as follows. Plainly, $\{F, E\} \Rightarrow F$. So by the deduction principle (23) for English, $F \Rightarrow$*if-*E*-then-*F. [To apply (23), we take $A, C = F$, $B = E$.] Thus, the truth of F guarantees the truth of the conditional *if-*E*-then-*F. Hence, (32)a is true. Next, consider (32)c. By the contradiction principle (24) for English, $\{\neg E, E\} \Rightarrow F$. So by the deduction principle (23) again, $\neg E \Rightarrow$*if-*E*-then-*F. Thus, the truth of $\neg E$ guarantees the truth of *if-*E*-then-*F. In other words, the falsity of E guarantees the truth of *if-*E*-then-*F. Hence, (32)c is true. Finally, consider (32)b. Suppose that E is true and F is false. Then E*-and-not-*F is true. Suppose for a contradiction that *if-*E*-then-*F is also true. Then by the first conditional principle (25) for English, *not-*$(E$*-and-not-*$F)$. Since it can't be the case that both E*-and-not-*F and *not-*$(E$*-and-not-*$F)$ are true, it must be that *if-*E*-then-*F is false. This establishes that if E is true and F is false then *if-*E*-then-*F is false; in other words, we've established (32)b.

So you see? The $\rightarrow$ of Sentential Logic represents *if–then–* of English [because (32) is true, hence (33) is true].[14]

8.4.3 Second argument showing that indicative conditionals are faithfully represented by $\rightarrow$

The next argument in favor of representing *if–then–* by $\rightarrow$ is drawn from Stalnaker [95]. Take two sentences E and F, and consider the complex sentence $($*not-*$E)$*-or-*F. We mean by the latter expression something like:

(36) Either it is not the case that E or it is the case that F (or maybe both of these possibilities hold).

[14] So far as we know, this argument is due to Mendelson [74]. (It appears in early editions of his book.)

This sentence ain't pretty but it is English, and the conditions under which (36) is true seem pretty clear, namely:

(37) if E is true and F is true then sentence (36) is true.
if E is true and F is false then sentence (36) is false.
if E is false and F is true then sentence (36) is true.
if E is false and F is false then sentence (36) is true.

Next, notice that (37) defines a truth table. The table specifies that (36) is true if E is true and F is true, etc. What do you notice about this truth table? Correct! It is the same truth table as the one for $E \rightarrow F$. It's no coincidence, of course, that the truth table for (36) is the same as the one for $E \rightarrow F$. Sentence (36) is nicely represented by $\neg E \vee F$ in $\mathcal{L}$, and $\neg E \vee F$ and $E \rightarrow F$ are logically equivalent.

Now suppose that we could show the following.

(38) (a) *if*-E-*then*-$F \Rightarrow$ (*not*-E)-*or*-F.
(b) (*not*-E)-*or*-$F \Rightarrow$ *if*-E-*then*-F.

Then [keeping in mind that (*not*-E)-*or*-F is Sentence (36)] we would know that the circumstances that render *if*-E-*then*-F true also make (36) true [this is (38)a], and also that the circumstances that render (36) true also make *if*-E-*then*-F true [this is (38)b]. In other words, we would know that *if*-E-*then*-F has the truth table shown in (37), which happens to be the truth table for $E \rightarrow F$. Hence, we will have shown that *if*-E-*then*-F is suitably represented in $\mathcal{L}$ by $E \rightarrow F$.

Let's show (38)a. By the first conditional principle (25),

if-E-*then*-$F \Rightarrow$*not*-(E-*and*-*not*-F).

By the DeMorgan principle (28),

not-(E-*and*-*not*-F) $\Rightarrow$(*not*-E)-*or*-*not*-*not*-F

[in (28), take $A = E$, $B =$ *not*-F]. So by the Transitivity principle (22),

$$\textit{if-}E\textit{-then-}F \Rightarrow (\textit{not-}E)\textit{-or-not-not-}F.$$

By the double negation principle (29),

$$(\textit{not-}E)\textit{-or-not-not-}F \Rightarrow (\textit{not-}E)\textit{-or-}F$$

so by transitivity again,

$$\textit{if-}E\textit{-then-}F \Rightarrow (\textit{not-}E)\textit{-or-}F.$$

That's (38)a. It's even easier to show (38)b. By Principal (27),

$$(\textit{not-}E)\textit{-or-}F \Rightarrow \textit{if-not-not-}E\textit{-then-}F$$

[taking $A =$ *not*-E, $B = F$ in (27)]. By the double negation principle (29),

$$\textit{if-not-not-}E\textit{-then-}F \Rightarrow \textit{if-}E\textit{-then-}F.$$

So by transitivity,

$$(\textit{not-}E)\textit{-or-}F \Rightarrow \textit{if-}E\textit{-then-}F,$$

which is (38)b.

So once again we see that the conditions under which *if*-E-*then*-F are true are exactly the conditions that make $E \rightarrow F$ true. Hence, $E \rightarrow F$ is a suitable representation of *if*-E-*then*-F in $\mathcal{L}$.

8.4.4 Could *if–then–* be truth-functional?

We hope that you are convinced that *if–then–* is nicely represented by $\rightarrow$ in $\mathcal{L}$. We shall also try to convince you that *if–then– cannot* be represented by $\rightarrow$. Before descending into this contradiction, however, let us discuss one reason you might have for doubting, even at this stage, that *if–then–* is represented by $\rightarrow$.

Assume (for the sake of argument) that *if–then–* is successfully represented by $\rightarrow$. Then every valid argument involving just $\rightarrow$ corresponds to a valid argument with just *if–then–*. Here are two such arguments.

(39) (a) $p \models (q \rightarrow p)$

(b) $\neg q \models (q \rightarrow p)$

They correspond to:

(40) (a) $p \Rightarrow$ *if-*q*-then-*p

(b) *not-*$q \Rightarrow$ *if-*q*-then-*p

Actually, (40)b involves more than just *if–then–* since negation is also present. But we have in mind such straightforward use of English negation that $\neg$ in (39)b can be trusted to represent it. So if $\rightarrow$ represents *if–then–*, you ought to agree that the arguments exhibited in (40) are secure. Now let p, q be as follows:

(41)

p	:	There is a one pound gold nugget on Mars.
q	:	Julius Caesar visited Sicily.

Then you ought to agree that the following arguments are secure.

(42) (a) There is a one pound gold nugget on Mars. Hence, if Julius Caesar visited Sicily, there is a one pound gold nugget on Mars.

(b) Julius Caesar did not visit Sicily. Hence, if Julius Caesar visited Sicily, there is a one pound gold nugget on Mars.

We expect that you'll find the arguments in (42) to be quirky at best. Should such inferences really be counted as secure?

As promised, we shall shortly present reasons to doubt that → adequately represents *if–then–* in $\mathcal{L}$. But we don't think that the odd quality of the arguments in (42) is one such reason. The oddness stems in part from the unrelatedness of p and q. Argument (42)a, for example, would sound better if (for example) p affirmed that Gaius Octavius (who became the emperor Augustus) visited Sicily. But the relatedness of sentences is a quixotic affair, varying with the background knowledge of the reasoner. If you were entertaining the hypothesis that wealthy aliens living in Sicily had invited Caesar there, and also left behind gold nuggets on Mars, then (41) would seem more connected. In assessing the security of inferences let us therefore leave aside issues of thematic integrity and any other consideration of whether the inference is likely to be useful in normal discourse. Such considerations belong more to the study of *pragmatics* than to logic.[15]

Pragmatics left to one side, the inferences in (42) might in fact be secure. Take the first one. If there really is a one pound gold nugget on Mars then how could the conclusion of the argument (the indicative conditional) turn out to be false? It couldn't turn out that Caesar did visit Sicily but Mars is nugget free, could it? We are therefore inclined to accept (42)a as secure.

Regarding (42)b, suppose that Julius Caesar did not visit Sicily. Then if he *did* visit Sicily we live in a contradictory world, and we've already agreed [in Principle (24)] that *every* sentence follows from a contradiction. Application of Principle (23) to $\{\textit{not-}p, p\} \Rightarrow q$ then yields (42)b.

Relegating the oddity of the inferences in (42) to pragmatics opens the door to embracing the *truth-functionality* of *if–then–*. To appreciate the issue, recall the truth tables for our connectives, given in Section 4.2.4. They specify the truth value of a larger formula like $\varphi \to \psi$ entirely in terms of the truth values of φ and ψ. Any other formula φ' could be substituted for φ in $\varphi \to \psi$ without changing the latter's truth value in a given truth-assignment α provided that α assigns the same truth value to φ and φ'. We discussed all this in the context of

[15] The *Encarta World English Dictionary* has the following entry for *pragmatics:* "the branch of linguistics that studies language use rather than language structure. Pragmatics studies how people choose what to say from the range of possibilities their language allows them, and the effect their choices have on those to whom they are speaking."

Fact (12) in Section 4.2.3. It follows that the truth of a formula like $\varphi \rightarrow \psi$ (we mean, the truth in Reality, the truth-assignment corresponding to the "real" world) is determined by nothing more than the truth values of φ and ψ. Now, if $\rightarrow$ successfully represents *if–then–* then we can expect the same to be true of indicative conditionals. Their truth value will depend on no more than the truth values of their left hand side and right hand side. Is this claim plausible? Consider the three sentences:

(43)
p : The Italian government will fall before 2007.
q : Elections will be held in Italy before 2007.
r : There is no Chinese restaurant in Twin Forks, Wyoming.

Suppose that p and r have the same truth value (both true or both false). Then the following two sentences will have the same truth value.

(44) If the Italian government falls before 2007 then elections will be held in Italy before 2007.

If there is no Chinese restaurant in Twin Forks, Wyoming, then elections will be held in Italy before 2007.

If this equivalence seems strange, we suggest attributing it to the pragmatic fact that Chinese restaurants and Italian politics are seldom juxtaposed in everyday discourse. The truth functionality of *if–then–* will then not be thrown into doubt by (44).

The idea that *if–then–* may be truth functional is all the more striking inasmuch as many other sentential connectives of English do not share this property. Consider the word "because." It also unites two sentences, just like *if–then–*. Even supposing that p and r of (43) share the same truth value, however, the following sentences appear not to.

Elections will be held in Italy before 2007 because the Italian government falls before 2007.

Elections will be held in Italy before 2007 because there is no Chinese restaurant in Twin Forks, Wyoming.

To be sure, it is not always easy to justify the distinction between pragmatic facts and purely semantic ones. Consider again argument (5), repeated here for convenience.

(45) John ran the marathon and died. Therefore, John died and ran the marathon.

There is definitely something peculiar about (45) but this may reflect nothing more than the pragmatic fact that people are expected to utter conjuncts in the temporal order in which they occur (a kind of story-telling). In this case, (45) counts as secure. Alternatively, the English word "and" might be polysemous, like *rocker*.[16] In particular, one of the meanings of "and" might be synonymous with "and then." If such is the sense of "and" in (45) then of course the inference is not secure (and not naturally translated by $p \wedge q \ / \ q \wedge p$). As noted earlier, which of these explanations is correct is the subject of much debate (see [33]). The present authors draw the pragmatics/semantics distinction where it seems to yield the cleanest overall theory of language. But you, as reader, will need to remain vigilant, and note disagreements that affect our claims about conditionals. For the moment, we have discounted worries about the inferences (42), and affirmed that *if–then–*, like →, is truth functional. So all looks swell for the idea of representing *if–then–* by → in $\mathcal{L}$. Our next task is to make this idea look not-so-swell.

8.5 Indicative conditionals cannot be represented by →

8.5.1 Our strategy, and more fiddling with the class of indicative conditionals

In this section we exhibit valid arguments in $\mathcal{L}$ involving → whose counterparts in English are not secure. That → is a poor representation of *if–then–* is demonstrated thereby. Other connectives figure in the arguments, notably

[16] Here is the definition of *polysemy*: "having multiple meanings; the existence of several meanings for a single word or phrase." The word *rocker* is polysemous because it can mean (*inter alia*) either a type of chair or a type of singer.

negation, conjunction, and disjunction. But it will be clear that the translation failure is due to the use of $\rightarrow$ to represent *if–then–*, rather than, e.g., the use of $\wedge$ to represent "and."

Actually, instead of writing particular valid arguments in $\mathcal{L}$, we'll use meta-variables like φ and ψ to describe entire classes of arguments that are valid in $\mathcal{L}$. Then we'll exhibit a translation of the schema into a non-secure English argument. We'll call the English argument a "counterexample" to the schema. This is enough to show that Criterion (20) is not satisfied.

The opposite strategy is not pursued. That is, we don't attempt to exhibit an invalid argument of $\mathcal{L}$ involving $\rightarrow$, all of whose translations into English are secure. Two examples of this latter kind are offered in (88) of Section 10.4.3, below. But here it will be simpler to stick to valid arguments in $\mathcal{L}$ with non-secure translation into English ("counterexamples" in the sense just defined). That'll be enough to make the point.

Even within our chosen strategy, we do not wish to exploit examples that rely on logical relations between the variables appearing in conditionals. To appreciate the issue, consider the following valid inference in $\mathcal{L}$.

(46) $\neg p \rightarrow q \models \neg q \rightarrow p$.

Choose p and q as follows.

p : Bob lives in Boston.
q : Bob lives somewhere in New England.

If we use $\rightarrow$ to represent *if–then–* then the validity (46) translates the non-secure argument:

(47) If Bob doesn't live in Boston then he lives somewhere in New England. Therefore, if Bob doesn't live in New England then he lives in Boston.

Argument (47) is not secure. Indeed, whereas the premise may well be true (if Bob lives, for example, in Worcester), the conclusion is *surely* false (since

Boston is in New England, duh ...).[17] We're not inclined to bend (47) to our present purposes, however, because we suspect trickery and can't identify the trick! Perhaps the example rests on the semantic connection between p and q, namely, the impossibility that p is true but q false. Since p and q are variables, this semantic connection cannot be represented in Sentential Logic. It is consequently unclear how to translate (47) into $\mathcal{L}$.[18]

To steer clear of such mysterious cases, let us therefore adjust once more the class of indicative conditionals. We agree to consider only indicative conditionals whose atomic constituents are *logically independent* of each other. In general, we call sentences $A_1 \ldots A_n$ "logically independent" just in case all combinations of truth and falsity among $A_1 \ldots A_n$ are possible (A_1 can be true and the other A_i false, etc.). In (47) it is thus required that the truth of "Bob lives in Boston" and the falsity of "Bob lives somewhere in New England" be jointly possible. Since this is not the case, we withdraw (47) from the class of indicative conditionals that can serve as counterexamples to our theories. This new limitation protects $\rightarrow$ from the invalid argument (47). But we'll now see that there are plenty of other cases in which $\rightarrow$ seems to misrepresent secure inferences involving *if–then–*.

8.5.2 Transitivity

Here is a principle from Sentential Logic whose validity is easy to check.

(48) FACT: $\{\varphi \rightarrow \psi, \psi \rightarrow \chi\} \models \varphi \rightarrow \chi$

[17] The example originates in Jackson [51], and is discussed in Sanford [89, pp. 138, 230].

[18] Don't be tempted to use $\models$ to code semantic relations among variables; $\models$ is not part of $\mathcal{L}$, but only an extension of English that allows us to talk about $\mathcal{L}$. See the remarks in Section 5.1.2.

That is, $\rightarrow$ has a *transitive* character.[19] For a counter-example, choose φ, ψ, and χ as follows.[20]

φ : The sun explodes tomorrow.
ψ : Queen Elizabeth dies tomorrow.
χ : There will be a state funeral in London within the week.

These choices yield the non-secure argument:

(49) COUNTEREXAMPLE: If the sun explodes tomorrow then Queen Elizabeth will die tomorrow. If Queen Elizabeth dies tomorrow, there will be a state funeral in London within the week. So, if the sun explodes tomorrow, there will be a state funeral in London within the week.

It seems quite possible for both premises of this argument to be true whereas the conclusion is certainly false. The argument is consequently not secure, which indicates that use of $\rightarrow$ to represent *if–then–* cannot be counted on to translate non-secure arguments of English into invalid arguments of Sentential Logic.

Are you having doubts? Witnessing the havoc wreaked by Argument (49), perhaps you're unwilling to declare it non-secure. But it won't be easy to defend the argument. Doesn't it seem just plain true — given the world the way it really is — that if the Queen dies tomorrow then she'll be honored with a state funeral shortly? We don't mean to claim that the foregoing conditional is somehow *necessarily* true; we agree that it is a *possibility* that the poor Queen

[19] In general, a relation (like *less than*) over a set of objects (like numbers) is said to be "transitive" just in case the relation holds between the objects x and z (in that order) if it holds between x and y and between y and z. In (48), the "objects" are formulas, φ, ψ and the relation is something like: "when $\rightarrow$ is inserted between φ, ψ, in that order, the resulting formula is true."

[20] This example was communicated to us by Paul Horwich many years ago. Another example appears in Adams [3], cited in Sainsbury [88, p. 76], namely:

> If Smith dies before the election then Jones will win. If Jones wins then Smith will retire from public life after the election. Therefore, if Smith dies before the election then he will retire from public life after the election.

could die without anyone noticing. We just mean that the second premise of (49) is a fact about contemporary society, or is likely to be so. The first premise is also factually correct (albeit not a necessary fact; the Queen might have made extraordinary contingency plans). It *could* therefore be the case that both premises of (49) are true. Yet the conclusion seems indubitably false (not necessarily, just in fact). Hence (49) is not secure. It may well lead from true premises to false conclusion. There is certainly no *guarantee* that its conclusion will be true if the premises are. Hence, the argument is insecure. Are you convinced? If not, don't worry (yet). There are more counterexamples coming your way.

8.5.3 Monotonicity

You can easily verify the validity of the following schema, often referred to as *monotonicity* or *left side strengthening*.

(50) FACT: $\varphi \rightarrow \psi \models (\varphi \wedge \chi) \rightarrow \psi$.

Counterexamples to (50) have been on offer for many years (see, for example Adams [3], Harper [44, p. 6], Sanford [89, p. 110]). Here's a typical example, based on the following choices of φ, ψ, χ.

φ	:	A torch is set to this book today at midnight.
ψ	:	This book is plunged into the ocean tonight at one second past midnight.
χ	:	This book will be reduced to ashes by tomorrow morning.

(51) COUNTEREXAMPLE: If a torch is set to this book today at midnight then it will be reduced to ashes by tomorrow morning. Therefore, if a torch is set to this book today at midnight and the book is plunged into the ocean tonight at one second past midnight then it will be reduced to ashes by tomorrow morning.

Given the way things actually are (namely, the book we're holding is dry, and we are far from the ocean), the premise of (51) is true; the darn thing really will be reduced to cinders if (God forbid) it is torched at midnight. At the same time, the conclusion of (51) isn't true (try it). Thus, the valid schema (50) translates into an non-secure argument if we represent *if–then–* by $\rightarrow$.

8.5.4 One way or the other

The following principle came to light just as logic began to take its modern form (see [89, p. 53]).

(52) FACT: $\models (\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$.

To render $(\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$ false, a truth-assignment would have to make φ true and ψ false, and also make ψ true and φ false. Fact (52) follows from the impossibility of such a truth-assignment.

Translation of (52) into English yields counterintuitive results. Pick at random some girl born in 1850, and consider:

φ : The girl grew up in Naples.
ψ : The girl spoke fluent Eskimo.

Then (52) yields:

(53) COUNTEREXAMPLE: At least one of the following statements is true.

If the girl grew up in Naples then she spoke fluent Eskimo.

If the girl spoke fluent Eskimo then she grew up in Naples.

We've here translated $\vee$ by "at least one," but the example seems just as forceful if "either–or–" is used instead. Since both of the conditionals in the disjunction seem false, (53) seems very unlike the tautology registered in (52). If you accept this judgment then you must doubt that *if–then–* is successfully represented by $\rightarrow$.[21]

[21] In place of (53) we were tempted by the following counterexample to (52).

8.5.5 Negating conditionals

We've saved the most convincing demonstration in this series for last. You can easily verify:

(54) FACT: $\neg(\varphi \rightarrow \psi) \models \varphi$.

The following kind of counterexample comes from Stevenson [96]. Let:

φ : God exists.
ψ : Evil acts are rewarded in Heaven.

So if $\rightarrow$ successfully represents *if–then–*, the following argument should be secure.

(55) COUNTEREXAMPLE: It is not true that if God exists then evil acts are rewarded in Heaven. Therefore, God exists.

Whatever your religious convictions (and we don't dare ask in today's political climate), surely (55) is a clunker when it comes to proving the existence of God. If you think otherwise, then you must also believe in Santa Claus. For, consider:

(56) COUNTEREXAMPLE: It is not true that if Santa exists then all good boys get lumps of coal for Christmas. Therefore, Santa exists.

Look, we're *sure* that the premise of (56) is true since if Santa exists *we* would never have been stuck with coal for Christmas (that would be totally un-Santa-like). But we're not committed thereby to believing that the old geezer actually exists. If you feel the same way, then (56) is another reason to doubt that $\rightarrow$ successfully represents *if–then–*.

(57) EXERCISE: Show that $(\varphi \wedge \psi) \rightarrow \chi \models (\varphi \rightarrow \chi) \vee (\psi \rightarrow \chi)$. Can you think of a counterexample to this principle?

If today is Monday then today is Tuesday, *or* if today is Tuesday then today is Monday.

What's wrong with this example?

8.6 Road map

We hope you are convinced by the arguments of the last section. We aimed to give you powerful reasons to doubt that *if–then–* is successfully represented by $\rightarrow$. Hence, you should doubt that the security of arguments in English is mirrored by their validity in Sentential Logic when $\rightarrow$ stands in for *if–then–*. At the same time, we hope that Section 8.4 gave you powerful reasons to agree that *if–then–* is, after all, successfully represented by $\rightarrow$. That is, you should accept that the security of arguments in English is mirrored by their validity in Sentential Logic when $\rightarrow$ stands in for *if–then–*.

If you are persuaded on both counts then you should now be experiencing *Existential Torment* (as if you discovered that Ronald Reagan was a commie double-agent). But don't despair; we're here to rescue you! In Chapter 10 we'll attempt to identify a false assumption that underlies *both* sides of the dilemma. Its denial allows us to resist the contradictory conclusions reached above. We can then consider afresh the question of representing *if–then–* in logic, and go on to develop an alternative account of indicative conditionals in English. This enterprise requires presenting some more logic, however. Specifically, in the next chapter we'll present the bare rudiments of *inductive logic* in the form of elementary probability theory. Our language $\mathcal{L}$ will still occupy center stage since the principal task will be to show how to assign probabilities to formulas in a sensible way. The new apparatus will then allow us to formulate yet another objection to representing *if–then–* via $\rightarrow$, and to throw into focus some of the assumptions that underlie our discussion so far. It will be quite a story, and you won't want to miss it! See you in Chapter 9.

Chapter 9

Probability in a Sentential Language

9.1 From truth to belief

As so far developed in this book, the fundamental idea of Sentential Logic is *truth*. Thus, the meaning of a formula is the set of truth-assignments that make it true, and validity is a guarantee of the truth of an argument's conclusion assuming the truth of its premises. The soundness and completeness theorems of Chapter 7 show that derivability is coincident with validity, hence another guide to the truth of conclusions given the truth of their premises. In the present chapter, we introduce a new idea into Sentential Logic, namely, *belief*. Specifically, we shall consider the logic of the "degree of confidence" or "firmness of belief" that a person may invest in a statement. The statements will all be formulas of our language $\mathcal{L}$ of sentential logic, introduced in Chapter 3. Degrees of confidence will be called *probabilities*.

In Section 1.3 we said that logic is the study of *secure inference*.[1] Chapter 5 examined the idea of secure inference from a semantic point of view, and Chapter 6 studied it from a syntactic or derivational point of view. In the present chapter, we take a new perspective, one that befits the transition from *deductive* to *inductive* logic. Inferences will now bear on belief rather than truth directly. They'll have the form "if the probabilities of formulas $\varphi_1 \ldots \varphi_n$ are such-and-such then the probability of formula ψ is so-and-so." Verifying such inferences will require thinking once again about truth-assignments, but now they will be conceived formally as *outcomes* in a *sample space*. To make this clear, the next section reviews elementary concepts of probability, apart from considerations of logic. Then we'll turn to probabilities for formulas of $\mathcal{L}$.

9.2 Probability Concepts

The little bit of standard probability that we need is easy. Beyond the basics, probability gets quite complicated. For an excellent introduction, see Ross [85].

[1] More exactly, logic is the study of several things at once, among them secure inference.

9.2.1 Sample spaces, outcomes, and events

Everything starts with a non-empty, finite set S, called a *sample space*. Insisting that S be finite avoids many technicalities and is sufficient for our purposes. Members of S are called *outcomes*. The idea is that S holds all the potential results from some "experiment." The experiment may be conducted artificially (by a person) or by Nature. Things are set up in such a way that the experiment will yield exactly one member of S. (You don't typically know in advance which one will happen.) For example, S might consist of the 30 teams in Major League Baseball, and the "experimental result" might be the success of just one team in the 2004 World Series. (If you read this after March 2004, please substitute an appropriate year, and the right number of teams.) There are 30 outcomes, namely, the Yankees, the Dodgers, etc. This kind of example is often an idealization. Thus, we ignore the "outcome" of no World Series in 2004 (e.g., because of a players' strike), just as we ignore the possibility that a tossed coin lands on its edge.

An *event* (over a sample space S) is a subset of S. In our example, some events are as follows.

(1) (a) the set consisting of just the Dodgers, that is: {Dodgers}.

(b) the set consisting of the Yankees, the Mets, the Dodgers, and the Giants that is: {Yankees, Mets, Dodgers, Giants}.

(c) the National League Western Division, that is:

$$\{\text{Dodgers}, \text{Giants}, \text{Diamondbacks}, \text{Rockies}, \text{Padres}\}.$$

Think of an event E as the claim that the experiment results in a member of E. Thus, event (1)c amounts to the claim that the winner of the 2004 World Series is one of the five teams mentioned, in other words, the claim that the winner comes from the National League Western Division.

Keep in mind that an event is a set of outcomes, not a description of that set. Thus the set of all MLB teams that once played in New York City is the same event as (1)b, namely, {Yankees, Mets, Dodgers, Giants}. And this is the same event as the set of teams that ever attracted the slightest devotion from the authors of this text (as it turns out). Similarly, we might describe an event using operations like intersection. For example, the intersection of (1)b with (1)c is just the event $\{\text{Dodgers}, \text{Giants}\}$. The same event can be described as the union of $\{\text{Dodgers}\}$ and $\{\text{Giants}\}$ or other ways.[2]

Here is another technical point. Outcomes are not events since the latter are subsets of S whereas the former are its members. This is why we put the braces around "Dodgers" in (1)a. In practice, we allow some sloppiness, and often talk about single outcomes as single-member events; thus, we often understand Dodgers to mean $\{\text{Dodgers}\}$.

9.2.2 Number of events, informativeness

How many events are there? This is the same question as "How many subsets of S are there?" Recall from your study of sets that there are 2^n subsets of a set with n members.[3] We thus have:

(2) FACT: There are 2^n events in a sample space of n elements.

For the sample space of 30 baseball teams there are thus 2^{30} events, more than a billion of them. Among them are two trivial but noteworthy cases. The space S and the empty set $\emptyset$ are both subsets of S. The event S amounts to the claim that one of the teams will win the 2004 World Series, which is essentially guaranteed. The event $\emptyset$ amounts to the claim that none of the teams will win the 2004 World Series, which is essentially impossible.

If one event is properly included in another, it is natural to consider the smaller one as more informative than the larger.[4] For example, the event *National League Western Division* — that is, the event (1)c — is properly included

[2] For example, as the set of teams that broke the hearts of millions of New Yorkers by perfidious transfer to California.

[3] See Section 2.6. In brief, every member of S can be either in or out of a given subset. These binary choices are independent, and n of them must be made. This yields $2 \times 2 \times \cdots \times 2$ (n times) $= 2^n$ combinations.

[4] Reminder: Set A is properly included in set B — written $A \subset B$ — just in case every member of A is a member of B but not conversely. See Section 2.2.

in the event *National League*. And it is more informative to claim that the 2004 World Series winner will come from the National League Western Division than to claim that the winner will come from the National League (provided that the claim is true). If two events are such that neither is included in the other, assigning relative informativeness is more delicate. For example, you might think that it is more informative to claim that the winner will come from $\{\text{Diamondbacks}\}$ than from $\{\text{Brewers}, \text{Pirates}\}$ since there is just one team in the first event and two in the second. But since it would take a *miracle* for either the Brewers or Pirates to even scrape out a winning season, it might also be said that $\{\text{Brewers}, \text{Pirates}\}$ is more informative than $\{\text{Diamondbacks}\}$. For now, we only compare information between events that are ordered by proper subset. The least informative event is therefore S itself since every other event is properly included in it. And the most informative claims are the singleton sets like $\{\text{Diamondbacks}\}$ since no other set is properly included in them — except for the empty set which we don't count as informative since it corresponds to a claim that must be false. We made similar remarks about information when discussing meanings in Section 4.3.3.

9.2.3 Probability distributions

Recall that S denotes our sample space. A "probability distribution" over S is any assignment of numbers to the outcomes of S such that (a) each number is drawn from the interval $[0, 1]$ (hence, can be interpreted as a probability), and (b) all the assigned numbers sum to unity. This idea can be put succinctly as follows.

(3) DEFINITION: A *probability distribution* over S is any function $\boldsymbol{Pr} : S \to [0, 1]$ such that $\sum_{s \in S} \boldsymbol{Pr}(s) = 1$.[5]

The expression "probability distribution over S" is often abbreviated to just "distribution" (provided it is clear which set S we're talking about).

[5]The symbol $\sum_{s \in S} \boldsymbol{Pr}(s)$ can be read: The sum of the probabilities assigned to the members s of S.

We illustrate the definition by making our baseball example more compact. Let S be the set of National League teams (instead of all teams), and think of the experiment as determining which will win the pennant. Then one distribution can be represented as follows.

(4)

Braves	$\frac{1}{32}$	Expos	$\frac{1}{16}$	Marlins	$\frac{1}{16}$	Mets	$\frac{1}{8}$
Phillies	$\frac{1}{16}$	Cardinals	$\frac{1}{16}$	Astros	$\frac{1}{64}$	Pirates	$\frac{1}{16}$
Cubs	$\frac{1}{64}$	Brewers	$\frac{1}{32}$	Dodgers	$\frac{1}{16}$	Diamondbacks	$\frac{1}{4}$
Giants	$\frac{1}{16}$	Rockies	$\frac{1}{32}$	Padres	$\frac{1}{32}$	Reds	$\frac{1}{32}$

According to (4), the probability that the Braves win the pennant is $1/32$, the probability that the Expos win is $1/16$, etc. We write $Pr(\text{Braves}) = 1/32$, $Pr(\text{Expos}) = 1/16$, and so forth. The numbers sum to unity since one of the teams is bound to win. There are, of course, other distributions, indeed, a limitless supply of them. If all the probabilities are the same (thus, $1/16$ in our example), the distribution is said to be *uniform*. At the other extreme, if all of the probabilities are zero except for one (which must therefore be unity), the distribution is said to be *dogmatic*.

9.2.4 Personal probability

Now you'll surely ask us "Which distribution is right, and how can you tell?" This innocent question opens the door to a complex debate about the nature of probability. For introduction to the issues, see Hacking [39], Gustason [38, Ch. 7] or Neapolitan [75, Ch. 2]. In the present work, we adopt a *personalist* or *subjective* perspective, and think of probabilities as reflecting the personal opinions of an idealized ratiocinator (thinking agent). To give meaning to such numbers, we take (for example) the attribution of $1/4$ probability to the Diamondbacks winning the pennant to mean that the agent finds the following bet to be fair. The agent wins \$3 if the Diamondbacks succeed in the pennant race, and pays \$1 if the Diamondbacks fail. More generally, let a bet on a given outcome involve the possibility of winning W dollars and losing L dollars. Then ascribing probability p to the outcome is reflected in the feeling that the bet is fair just in case $p = L/(W + L)$. To see why it is plausible to find such a bet fair, let us define the *expectation* of a bet. Suppose that you stand to gain W dollars if an event E comes to pass, and lose L dollars otherwise. Suppose also that you assign probability p to E occurring. Then your expectation for this bet is:

(5) $[\ p \times W\ ] - [\ (1-p) \times L\ ].$

In other words, your expectation is the probability of winning times the gain you'll receive minus the probability of losing times the loss you sustain. We hope that it will strike you as obvious that for the bet to be fair, its expectation should be zero (then it favors neither party). For example, the preceding bet on the Diamondbacks is fair since

$$\$3 \times \frac{1}{4} \quad - \quad \$1 \times (1 - \frac{1}{4}) = 0.$$

You can now see why you should take a bet on E to be fair if the ratio $L/(W+L)$ of losses to wins-plus-losses equals your probability that the E will happen. If $p = \frac{L}{W+L}$ then the expression (5) resolves to:

$$[\ p \times W\ ] - [\ (1-p) \times L\ ]$$

$$= [\ \frac{L}{W+L} \times W\ ] - [\ (1 - \frac{L}{W+L}) \times L\ ]$$

$$= [\ \frac{L}{W+L} \times W\ ] - [\ (\frac{W}{W+L}) \times L\ ] = 0.$$

Such an approach to fairness gives quantitative form to the intuition that winnings should be higher when betting on an improbable event (or losses should be lower). This is because higher W (or lower L) is needed to balance smaller p if the expression in (5) is to equal zero.

If you find a given bet to be fair then you should be indifferent between which side you take, that is, whether you receive W with probability p or L with probability $(1-p)$. For example, if the bet about the Diamondbacks' winning the pennant is fair for you then it should not matter whether (a) you gain \$3 if the Diamondbacks succeed and lose \$1 otherwise, or (b) you win \$1 if the Diamondbacks fail and lose \$3 if they succeed.

Note that the fairness of a bet concerns a given individual, namely, the one whose probabilities are at issue. Another person with different probabilities may find the same bet to be biased (in one direction of the other), and thus find a different bet to be fair. Such relativity to a particular individual makes sense in our "personalistic" framework. Probabilities reflect opinions, which may vary across individuals. Invoking fair bets is intended only to give content to the idea that an individual assigns a particular probability to a particular outcome.[6]

Let us admit that this way of explaining probabilities is not entirely satisfactory. For one thing, you might *like* the Diamondbacks, and prefer betting in their favor rather than against them. This will distort the probabilities we attribute to you. You might also find losing a sum of money to be more painful than gaining the same amount (which may well be the case of most of us; see Tversky & Kahneman [57]). In this case, the relation between probability and (monetary) bets will again be distorted. For another difficulty, suppose you think that the probability of your becoming a multi-billionaire next week is only $.0001$, leading to a bet in which you win \$99,990 if you become a multi-billionaire next week and lose \$10 if you don't. It is not obvious that this bet is genuinely fair since the added \$99,990 is chicken feed to a multi-billionaire whereas you could really use the \$10 you risk losing next week. Despite these problems (often discussed in the literature on subjective probability) the idea of a fair bet should suffice to indicate the interpretation of probability adopted here.[7]

In a nutshell, probability reflects confidence, or its inverse, doubt. A person whose distribution over National League teams is uniform suffers the most doubt; every team is given the same chance of winning the pennant. If the distribution is dogmatic, there is no doubt at all; a single team has every chance to win, the others none. In between these extremes is every conceivable pattern of relative doubt and confidence. If the distribution is (4), for example, there

[6]For more on probability and bets, see Skyrms [91].

[7]For extended discussion, see Howson & Urbach [50, Ch. 5], Chihara [17].

is some confidence in the Diamondbacks ending up on top, but also plenty of doubt reflected in the nontrivial probabilities assigned to other teams.

9.2.5 Probabilities assigned to events

So far we've only considered the probability of outcomes, that is, members of the sample space S. How can we extend this idea to events over S? The natural thing to do is add up the probabilities of the outcomes that comprise the event. The matter can be put this way.[8]

(6) DEFINITION: Suppose that *Pr* is a probability distribution over the sample space S. We extend *Pr* to the set $\mathcal{E} = \{E \mid E \subseteq S\}$ of events over S. For $E \in \mathcal{E}$ we define: $\boldsymbol{Pr}(E) = \Sigma_{o \in E} \boldsymbol{Pr}(o)$.

Consider, for example, the distribution given in (4). What probability does it assign to

$$\{\text{Dodgers}, \text{Giants}, \text{Diamondbacks}, \text{Rockies}, \text{Padres}\},$$

namely, the event that the pennant winner comes from the National League Western Division? We see that:

$$\boldsymbol{Pr}(\text{Dodgers}) = 1/16 \quad \boldsymbol{Pr}(\text{Giants}) = 1/16 \quad \boldsymbol{Pr}(\text{Diamondbacks}) = 1/4$$
$$\boldsymbol{Pr}(\text{Rockies}) = 1/32 \qquad \boldsymbol{Pr}(\text{Padres}) = 1/32$$

Adding these numbers yields

$$\boldsymbol{Pr}(\{\text{Dodgers}, \text{Giants}, \text{Diamondbacks}, \text{Rockies}, \text{Padres}\}) = \frac{28}{64} = \frac{7}{16}.$$

It makes sense to add the probabilities of each outcome in a given event because the outcomes are mutually exclusive; if one occurs, no other does.

[8] *Notation*: In what follows, Σ represents summation. If $x_1, x_2, \dots x_n$ are n numbers then $\Sigma_{i \leq n} x_i$ is their sum, and $\Sigma_{i \leq n} x_i^2$ is the sum of their squares. The expression $\Sigma_{o \in E} \boldsymbol{Pr}(o)$ is the sum of the probabilities assigned to outcomes in the event E.

We consider Definition (6) to *extend* the function ***Pr*** introduced in Definition (3). This is because the two definitions agree about the probabilities assigned to outcomes in S, namely, they both give what ***Pr*** gave originally. But Definition (6) goes further by giving a value to ***Pr*** when it is applied to events. It was noted earlier that outcomes are sometimes conceived as events whose braces have been omitted. It is for this reason that our extended function ***Pr*** gives the same number to an outcome $x \in S$ as it does to the event $\{x\} \subseteq S$. Observe that $\boldsymbol{Pr}(\{x\}) = \Sigma_{o \in \{x\}} \boldsymbol{Pr}(o) = \boldsymbol{Pr}(x)$.

Let us return briefly to the "informativeness" of events, discussed in Section 9.2.2. In the context of a specific distribution ***Pr***, it is natural to consider event E_1 to be more informative than event E_2 if $\boldsymbol{Pr}(E_1) < \boldsymbol{Pr}(E_2)$. The idea is that we learn more when something surprising happens compared to something obvious; and surprising events have lower probabilities. To quantify the information in an event E (relative to a distribution ***Pr***), statisticians often use $-\log_2 \boldsymbol{Pr}(E)$ (since this expression has convenient properties). It's easy to see that $\boldsymbol{Pr}(E_1) < \boldsymbol{Pr}(E_2)$ if and only if $-\log_2 \boldsymbol{Pr}(E_1) > -\log_2 \boldsymbol{Pr}(E_2)$; that is, the formal definition of informativeness is inversely related to probability, as intended. For a more complete discussion, see [85, §9.3]..

9.2.6 Probabilities assigned to conditional events

We are not finished extending ***Pr***. We must also consider *conditional events* like "a seaport team will win the pennant *supposing that* some team in the National League Western Division does." Such conditional events are conceived as ordered pairs of ordinary events.[9] In the foregoing example, the pair is (E, F), where

(7)
$$\begin{array}{rcl} E & = & \{\text{Marlins, Mets, Dodgers, Giants, Astros, Padres}\} \\ F & = & \{\text{Dodgers, Giants, Diamondbacks, Rockies, Padres}\} \end{array}$$

It is customary to elongate the comma between the two events, making it into a bar, and to drop the outer parentheses. The conditional event in question is then denoted $E \mid F$. Our goal is to extend ***Pr*** to embrace such events, so that we

[9] You studied ordered pairs in Section 2.9.

can write $\textbf{\textit{Pr}}(E \mid F) = .1$ to express our conviction that *assuming* the pennant winner comes from the National League Western Division, it is unlikely to be a seaport team. We proceed as follows.[10]

(8) DEFINITION: Suppose that *Pr* is a probability distribution over the sample space S. We extend *Pr* to all pairs $E \mid F$ of events over S for which $\textbf{\textit{Pr}}(F) > 0$. Given any such pair $E \mid F$, we define:

$$\textbf{\textit{Pr}}(E \mid F) = \frac{\textbf{\textit{Pr}}(E \cap F)}{\textbf{\textit{Pr}}(F)}.$$

If $\textbf{\textit{Pr}}(F) = 0$ then $\textbf{\textit{Pr}}(E \mid F)$ is not defined.

$\textbf{\textit{Pr}}(E \mid F)$ is not defined if $\textbf{\textit{Pr}}(F) = 0$ for otherwise there would be division by zero. We illustrate Definition (8) with the events in (7) and the distribution (4). We see that $E \cap F = \{$Dodgers, Giants, Padres$\}$. So by (4) we have $\textbf{\textit{Pr}}(E \cap F) = \frac{5}{32}$ and $\textbf{\textit{Pr}}(F) = \frac{7}{16}$. Hence,

$$\textbf{\textit{Pr}}(E \mid F) = \frac{\textbf{\textit{Pr}}(E \cap F)}{\textbf{\textit{Pr}}(F)} = \frac{\frac{5}{32}}{\frac{7}{16}} = \frac{5}{14}.$$

Let it be emphasized that the probability of an event as well as the probability of a conditional event depend on the underlying distribution *Pr*. Different choices of distribution at the outset yield different probabilities of events and conditional events.

Events that are not conditional are known as *absolute*. In our example, both E and F are absolute (in contrast to $E \mid F$, which is conditional). It is also said that $\textbf{\textit{Pr}}(E)$ is an "absolute probability" whereas $\textbf{\textit{Pr}}(E \mid F)$ is a "conditional probability."[11]

[10] Recall from Section 2.4 that $A \cap B$ denotes the intersection of the sets A and B, that is, the set of elements common to A and B.

[11] Another common terminology (e.g., in Cohen [20]) is to call absolute probabilities *monadic* and conditional probabilities *dyadic*.

9.2.7 Conditional versus absolute probability

Conditional probabilities are attached to pairs of events, instead of to single events. Is this complication really necessary? Perhaps for each pair of events there is a single event that expresses what the pair expresses. It is not altogether clear how a single event E could express what a pair $F \mid G$ expresses, but at minimum the following would be true.

(9) For all probability distributions ***Pr*** for which $\boldsymbol{Pr}(G) > 0$, $\boldsymbol{Pr}(E) = \boldsymbol{Pr}(F \mid G)$.

If (9) holds then the absolute probability of E is the conditional probability of $F \mid G$ with respect to any distribution in which the latter probability is defined. Conditional probability would then be dispensable in the sense that we could replace conditional events with absolute events of equal probability.

Notice that the idea of replacing each conditional event $F \mid G$ with an absolute event E amounts to defining a function f that maps conditional events into absolute events. We write $f(F \mid G) = E$ to mark the use of E to replace $F \mid G$. Since conditional events are just pairs drawn from the set $\mathcal{E}$ of all events, we see that such a function f has domain $\mathcal{E} \times \mathcal{E}$ and range $\mathcal{E}$.[12]

It turns out that conditional probabilities are *not* dispensable. There is no way to match pairs of events with single events such that the conditional probability of the former is the absolute probability of the latter. The matter can be stated precisely as follows.

(10) THEOREM: Suppose that the sample space S includes at least three outcomes. Then there is no function $f : \mathcal{E} \times \mathcal{E} \rightarrow \mathcal{E}$ such that for all $e_1, e_2 \in \mathcal{E}$ and probability distributions ***Pr*** with $\boldsymbol{Pr}(e_2) > 0$, $\boldsymbol{Pr}(e_1 \mid e_2) = \boldsymbol{Pr}(f(e_1, e_2))$.

Let us state the theorem another way. We're considering a sample space S with at least three outcomes (like in the baseball examples above). Choose

[12] For the $\times$ notation, see Definition (23) in Section 2.9. Functions were introduced in Section 2.10.

any function f that maps pairs of events from the sample space into single events. Thus, given two events $e_1, e_2 \subseteq S$, $f(e_1, e_2)$ is a subset of S, that is, an event from S. Now choose a probability distribution ***Pr***. We are hoping that for any two events $e_1, e_2 \subseteq S$, if $\boldsymbol{Pr}(e_2) > 0$ (so that the conditional probability of $e_1 \mid e_2$ is defined when using ***Pr***), $\boldsymbol{Pr}(e_1 \mid e_2) = \boldsymbol{Pr}(f(e_1, e_2))$. Alas, no matter what function f we choose, the latter equality will sometimes be false. The theorem reformulates a result due to David Lewis [67]. Our proof is an adaptation of Bradley [13]. It's OK to skip it; just rejoin the discussion in Section 9.2.8, below.

PROOF OF THEOREM (10): Suppose that S includes at least three outcomes, o_1, o_2, o_3. Choose any function $f : \mathcal{E} \times \mathcal{E} \to \mathcal{E}$. Let events $a = \{o_1, o_2\}$ and $b = \{o_2, o_3\}$ be given. It suffices to show:

(11) For some distribution ***Pr***,

$$\boldsymbol{Pr}(b) > 0 \text{ and } \boldsymbol{Pr}(a \mid b) = \frac{\boldsymbol{Pr}(a \cap b)}{\boldsymbol{Pr}(b)} \neq \boldsymbol{Pr}(f(a, b)).$$

We distinguish two cases depending on whether $f(a, b) \subseteq b$.

Case 1: $f(a, b) \subseteq b$. Choose a distribution ***Pr*** such that $\boldsymbol{Pr}(o_1) > 0$, $\boldsymbol{Pr}(o_2) > 0$ and $\boldsymbol{Pr}(o_3) = 0$. Then the choice of a and b implies that $0 < \boldsymbol{Pr}(a \cap b) = \boldsymbol{Pr}(b) < 1$. Hence:

(12) $$\frac{\boldsymbol{Pr}(a \cap b)}{\boldsymbol{Pr}(b)} = 1.$$

Since $f(a, b) \subseteq b$ (the present case), $\boldsymbol{Pr}(f(a, b)) \leq \boldsymbol{Pr}(b) < 1$. Hence, $\boldsymbol{Pr}(f(a, b)) < 1$. Also, since $\boldsymbol{Pr}(o_2) > 0$, $\boldsymbol{Pr}(b) > 0$. The latter facts in conjunction with (12) imply (11).

Case 2: $f(a, b) \not\subseteq b$. Then $f(a, b) \cap \overline{b} \neq \emptyset$.[13] Choose $o^* \in f(a, b) \cap \overline{b}$. We have $o^* \neq o_2$ (since $o^* \notin b$) so we may choose a distribution ***Pr*** such that $\boldsymbol{Pr}(o_2) = 0$, $\boldsymbol{Pr}(o^*) > 0$, and $\boldsymbol{Pr}(o_3) > 0$. Then:

(13) (a) $\boldsymbol{Pr}(f(a, b)) > 0$ [because $o^* \in f(a, b)$ and $\boldsymbol{Pr}(o^*) > 0$].

[13] $\overline{b}$ denotes the complement of b in S. See Section 2.3.

(b) $\mathbf{Pr}(b) > 0$ [because $o_3 \in b$ and $\mathbf{Pr}(o_3) > 0$].

(c) $\mathbf{Pr}(a \cap b) = 0$ [because $a \cap b = \{o_2\}$ and $\mathbf{Pr}(o_2) = 0$].

From (13)b,c,

$$\frac{\mathbf{Pr}(a \cap b)}{\mathbf{Pr}(b)} = 0.$$

In conjunction with (13)a, the latter fact implies (11). ∎

9.2.8 Changing distributions

Nothing lasts forever, and our beliefs, in particular, are usually in flux. What should you do if your probability for an event E increases to unity? Then you'll need to change your distribution from its original state, say $\mathbf{Pr}_1$, to some revised state, say $\mathbf{Pr}_2$.

For concreteness, suppose that the sample space consists of four outcomes a, b, c, d with (starting) probabilities $.1, .2, .3, .4$, respectively. Let this be the distribution $\mathbf{Pr}_1$. Suppose that E is the event $\{b, c\}$. Thus, $\mathbf{Pr}_1(E) = .5$. Imagine that your confidence in E now changes to certainty, perhaps because of some new experience, perhaps because you've reflected some more. So your new distribution, $\mathbf{Pr}_2$, should be such that $\mathbf{Pr}_2(E) = 1.0$. As a consequence, you must also change your probabilities for a, b, c, d since $\mathbf{Pr}_1(b)$ and $\mathbf{Pr}_1(c)$ don't sum to unity, as required by $\mathbf{Pr}_2$. How should you adjust the probabilities of a, b, c, d to transform $\mathbf{Pr}_1$ into $\mathbf{Pr}_2$?

The standard response is to set $\mathbf{Pr}_2(x) = \mathbf{Pr}_1(x \mid E)$ for each $x \in \{a, b, c, d\}$. In this case, we get:

$$\mathbf{Pr}_2(a) = \frac{\mathbf{Pr}_1(\{a\} \cap \{b, c\})}{\mathbf{Pr}_1(\{b, c\})} = \frac{\mathbf{Pr}_1(\emptyset)}{\mathbf{Pr}_1(\{b, c\})} = \frac{0}{.5} = 0.$$

$$\mathbf{Pr}_2(b) = \frac{\mathbf{Pr}_1(\{b\} \cap \{b, c\})}{\mathbf{Pr}_1(\{b, c\})} = \frac{\mathbf{Pr}_1(b)}{\mathbf{Pr}_1(\{b, c\})} = \frac{.2}{.5} = .4.$$

$$\boldsymbol{Pr}_2(c) = \frac{\boldsymbol{Pr}_1(\{c\} \cap \{b,c\})}{\boldsymbol{Pr}_1(\{b,c\})} = \frac{\boldsymbol{Pr}_1(c)}{\boldsymbol{Pr}_1(\{b,c\})} = \frac{.3}{.5} = .6.$$

$$\boldsymbol{Pr}_2(d) = \frac{\boldsymbol{Pr}_1(\{d\} \cap \{b,c\})}{\boldsymbol{Pr}_1(\{b,c\})} = \frac{\boldsymbol{Pr}_1(\emptyset)}{\boldsymbol{Pr}_1(\{b,c\})} = \frac{0}{.5} = 0.$$

Notice that $\boldsymbol{Pr}_2$ is a genuine distribution over $\{a, b, c, d\}$ since it sums to unity. It also gives the desired probability to E, namely, unity.

The foregoing advice for revising a distribution when an event comes to be endowed with certainty is known as the *conditionalization* doctrine. For its justification, see Resnik [83, Ch. 3-3d]. For extension of the doctrine to events whose probabilities change to values other than certainty, see Jeffrey [53, Ch. 11].

That's all you need from the elementary theory of probability. Now we show how to transfer these ideas to $\mathcal{L}$, our language of sentential logic.[14]

9.3 Probability for $\mathcal{L}$

Recall that we fixed the number of sentential variables in $\mathcal{L}$, once and for all, back in Section 3.2. We agreed to denote this number by n. For illustrations we'll assume, as usual, that $n = 3$.

9.3.1 Truth assignments as outcomes

To get our project off the ground, we need to identify the sample space relevant to $\mathcal{L}$. In our discussion of probability concepts [Section 9.2.1, above], any (finite) nonempty set S could serve as sample space. The elements of S were then conceived as potential results of an experiment that chooses one member of S as "outcome." To transfer these ideas to $\mathcal{L}$, we take the sample space to be the set of truth-assignments. Recall from Fact (4) in Section 4.2.1 that there are 2^n truth-assignments. And recall from Definition (5) in the same section that the

[14]More thorough treatments of the material that follows are available in [77, 41, 43]. For a history of these ideas, see [40].

set of all truth-assignments for $\mathcal{L}$ is denoted by TrAs. Hence, our sample space is TrAs, and outcomes are the individual truth-assignments that compose TrAs.

To make intuitive sense of this terminology we must regard truth-assignments as potential results of an experiment. The idea is to conceive each truth-assignment as one way the world might have turned out to be after Nature's choice of the "actual" world Reality from TrAs. Each truth-assignment is thus a potential outcome of Nature's selection. (This conception of truth-assignments was introduced in Section 4.3.1.)

9.3.2 Distributions over TrAs

Since distributions in the general setting are assignments of numbers to outcomes, distributions in the logical setting are assignments of numbers to truth-assignments. More precisely:

(14) DEFINITION: A *probability distribution for $\mathcal{L}$* is any function $\mathit{Pr} : \mathsf{TrAs} \rightarrow [0, 1]$ such that $\sum \mathit{Pr}(s) = 1$, where the sum is over all $s \in \mathsf{TrAs}$.

We usually abbreviate the expression "probability distribution for $\mathcal{L}$" to just "distribution." The following distributions illustrate the definition.

(15) (i)

	p	q	r	*prob*
(a)	t	t	t	.15
(b)	t	t	f	.1
(c)	t	f	t	0
(d)	t	f	f	.05
(e)	f	t	t	.25
(f)	f	t	f	.15
(g)	f	f	t	.1
(h)	f	f	f	.2

(ii)

	p	q	r	*prob*
(a)	t	t	t	1/8
(b)	t	t	f	1/8
(c)	t	f	t	1/8
(d)	t	f	f	1/8
(e)	f	t	t	1/8
(f)	f	t	f	1/8
(g)	f	f	t	1/8
(h)	f	f	f	1/8

(iii)

	p	q	r	*prob*
(a)	t	t	t	0
(b)	t	t	f	0
(c)	t	f	t	0
(d)	t	f	f	0
(e)	f	t	t	1
(f)	f	t	f	0
(g)	f	f	t	0
(h)	f	f	f	0

According to (15)*i*, the probability that all three variables are true is $.15$, and the probability that all three are false is $.2$. The "uniform" distribution (15)*ii* sets these two probabilities to $1/8$ (same as for the other truth-assignments), whereas the "dogmatic" distribution (15)*iii* sets them both to 0 (placing all confidence in the the fifth truth-assignment in the list).

9.3.3 Events and their probabilities

Since an event is a subset of the sample space, events in the logical context are subsets of TrAs, hence, sets of truth-assignments. The set $\{(a), (b), (c), (d)\}$ thus denotes an event. Recall from Section 4.3.2 that subsets of TrAs are also known as *meanings*. In Definition (25) of that section, the set of all meanings was given the name Meanings. Hence, Meanings is the set of all events.

Now recall from Definition (28) of Section 4.4.1 that each formula φ of $\mathcal{L}$ is associated with a meaning, denoted $[\varphi]$. The heart of the matter is the (unforgettable) equation:

(16) $[\varphi] = \{\alpha \in \mathsf{TrAs} \mid \alpha \models \varphi\}$.

For example, $[p] = \{(a), (b), (c), (d)\}$. The probability of a given event (meaning) is found by adding up the probabilities of its members. The probability of $\{(a), (b), (c), (d)\}$, for example, comes from adding up the probabilities of each of (a), (b), (c), and (d). So it is natural to take the probability of p to likewise be the sum of the probabilities of (a), (b), (c), and (d). In other words, we take the probability of $\varphi \in \mathcal{L}$ to be the probability of the event $[\varphi]$, the meaning of φ. Relying on Equation (16), we may express the matter this way.

(17) DEFINITION: Let distribution *Pr* be given. Then *Pr* is extended to $\mathcal{L}$ as follows. For all $\varphi \in \mathcal{L}$,

$$\boldsymbol{Pr}(\varphi) = \Sigma_{\alpha \models \varphi} \boldsymbol{Pr}(\alpha).$$

Perhaps you feel more comfortable writing the equation in (17) as follows.

$$\boldsymbol{Pr}(\varphi) = \Sigma_{\alpha \in [\varphi]} \boldsymbol{Pr}(\alpha).$$

The two equations are equivalent in view of (16).

We illustrate with the distribution (15)*i*, above. What is $\boldsymbol{Pr}(p)$? Well, $\boldsymbol{Pr}(p) = \sum\{\boldsymbol{Pr}(\alpha) \mid \alpha \models (p)\} = \boldsymbol{Pr}(a)+\boldsymbol{Pr}(b)+\boldsymbol{Pr}(c)+\boldsymbol{Pr}(d) = .15+.1+0+.05 = .30$. Hence, $\boldsymbol{Pr}(p) = .30$. What is $\boldsymbol{Pr}(p \vee \neg r)$? Well, $\boldsymbol{Pr}(p \vee \neg r) = \sum\{\boldsymbol{Pr}(\alpha) \mid \alpha \models (p \vee \neg r)\} = \boldsymbol{Pr}(a) + \boldsymbol{Pr}(b) + \boldsymbol{Pr}(c) + \boldsymbol{Pr}(d) + \boldsymbol{Pr}(f) + \boldsymbol{Pr}(h) = .15 + .1 + 0 + .05 + .15 + .2 = .65$. Hence, $\boldsymbol{Pr}(p \vee \neg r) = .65$.

Let us recall the following fact from Section 5.5.

(18) For every $M \subseteq \mathsf{Meanings}$ there is $\varphi \in \mathcal{L}$ such that $[\varphi] = M$.

For example, the set $\{(c), (d)\}$ is the meaning of $p \wedge \neg q$. The significance of (18) is that we can think in terms of the probability of formulas without fear of missing any events. The probability of $\{(c), (d)\}$, for example, is expressed by $\boldsymbol{Pr}(p \wedge \neg q)$. Indeed, Fact (65) of Section 5.7 tells us that infinitely many formulas express any given meaning. So the probability of $\{(c), (d)\}$ can be expressed using any of the infinitely many formulas that mean $\{(c), (d)\}$, e.g., $\neg q \wedge p$, $\neg(q \vee \neg p)$, etc.

(19) EXERCISE: According to the probability distribution (15)i, what are $\boldsymbol{Pr}(p \wedge r)$ and $\boldsymbol{Pr}(r \rightarrow \neg q)$?

9.3.4 Facts about probability

Let a probability distribution *Pr* be given. We list a bunch of facts about *Pr*. In each case, the proof is straightforward, and we'll just provide hints. Working through these facts is a great way of getting clear about probabilities in our language $\mathcal{L}$.

(20) FACT: For all $\varphi \in \mathcal{L}$, if $\models \varphi$ then $\boldsymbol{Pr}(\varphi) = 1$.

This is because $[\varphi] = \mathsf{TrAs}$ if $\models \varphi$. Fact (20) makes sense. Tautologies are certainly true (because vacuous). They should have probability 1.

(21) FACT: For all $\varphi, \psi \in \mathcal{L}$, if $\psi \models \varphi$ then $\boldsymbol{Pr}(\psi) \leq \boldsymbol{Pr}(\varphi)$.

This is because $[\psi] \subseteq [\varphi]$ if $\psi \models \varphi$. For example, $\boldsymbol{Pr}(\varphi \wedge \psi) \leq \boldsymbol{Pr}(\varphi)$.[15] If $\psi \models \varphi$ then ψ makes a claim that is at least as strong as the claim of φ. Stronger claims have greater chance of being false than weaker claims, which is what (21) expresses.

(22) FACT: For all $\varphi, \psi \in \mathcal{L}$, if $\models \varphi \leftrightarrow \psi$ then $\boldsymbol{Pr}(\varphi) = \boldsymbol{Pr}(\psi)$.

This is because $[\varphi] = [\psi]$ if $\models \varphi \leftrightarrow \psi$. To illustrate, $\boldsymbol{Pr}(p) = \boldsymbol{Pr}(p \vee (r \wedge \neg r))$. Logically equivalent formulas express the same meaning, so they ought to have the same probability.

(23) FACT: For all $\varphi \in \mathcal{L}$, $\boldsymbol{Pr}(\varphi) + \boldsymbol{Pr}(\neg\varphi) = 1$. [Hence, $\boldsymbol{Pr}(\neg\varphi) = 1 - \boldsymbol{Pr}(\varphi)$.]

This is because $[\neg\varphi] = \mathsf{TrAs} - [\varphi]$.

(24) FACT: If $\varphi \in \mathcal{L}$ is a contradiction, $\boldsymbol{Pr}(\varphi) = 0$.

[15] Simple though the latter principle may appear, ordinary intuition about chance often fails to honor it. See [103] and references cited there.

This is because $[\varphi] = \emptyset$ if φ is a contradiction. Contradictions can't be true. So they must have probability 0.

(25) FACT: (Law of total probability) For all $\varphi, \psi \in \mathcal{L}$, $\boldsymbol{Pr}(\varphi \wedge \psi) + \boldsymbol{Pr}(\varphi \wedge \neg\psi) = \boldsymbol{Pr}(\varphi)$.

This is because $[\varphi \wedge \psi] \cup [\varphi \wedge \neg\psi] = [\varphi]$, and $[\varphi \wedge \psi] \cap [\varphi \wedge \neg\psi] = \emptyset$. A more general version of the law may be stated as follows.

(26) FACT: Let $\varphi_1 \cdots \varphi_n \in \mathcal{L}$ be such that for all distinct $i, j \leq n$, $\varphi_i \models \neg\varphi_j$. Then $\boldsymbol{Pr}(\varphi_1 \vee \cdots \varphi_n) = \boldsymbol{Pr}(\varphi_1) + \cdots \boldsymbol{Pr}(\varphi_n)$.

The condition $\varphi_i \models \neg\varphi_j$ means that φ_i and φ_j are satisfied by different truth-assignments; they are never true together. For an example, consider the eight formulas:

(27)

$p \wedge q \wedge r$	$p \wedge q \wedge \neg r$	$p \wedge \neg q \wedge r$	$p \wedge \neg q \wedge \neg r$
$\neg p \wedge q \wedge r$	$\neg p \wedge q \wedge \neg r$	$\neg p \wedge \neg q \wedge r$	$\neg p \wedge \neg q \wedge \neg r$

Each is satisfied by exactly one of the eight truth-assignments over p, q, r, so each formula implies the negation of the others. It should be clear that the sum of the probabilities assigned to these formulas must be 1.0.

9.3.5 Necessary and sufficient conditions for probability

This section won't be used later. If you prefer to skip it, just pick us up in Section 9.3.6, below.

Consider any function $F : \mathcal{L} \rightarrow [0, 1]$ that maps each formula of $\mathcal{L}$ into a number between 0 and 1 (inclusive). Let's say that F *represents* a probability distribution just in case there is some distribution ***Pr*** (hence some function from TrAs to [0,1]) that yields F via Definition (17). We've seen that if F represents a probability distribution then F honors the properties recorded in Facts (20), (22), and (26), among others. In other words, these three properties are *necessary* for F to represent a probability distribution. They are also *sufficient*. To see this, let formulas $\psi_1 \cdots \psi_m$ be such that every truth-assignment satisfies exactly one of the ψ_i's, and every ψ_i is satisfied by exactly one truth-assignment. The conjunctions in (27) constitute such a list, for the case of three variables.[16] Observe that a truth-assignment satisfies a given formula just in case the corresponding ψ_i implies that formula. In light of Definition (17), for F to represent a distribution it suffices that:

(a) $F(\psi_1) + \cdots F(\psi_m) = 1$, and

(b) for all φ, $F(\varphi) = \sum_i F(\psi_i)$ for all $i \leq m$ such that $\psi_i \models \varphi$.

The first condition follows from (20), (26) and the fact that $\models \psi_1 \vee \cdots \vee \psi_m$. The second condition follows from (22), (26) and the fact that every formula is logically equivalent to a disjunction of some subset of the ψ_i's. The latter claim is an easy corollary of Corollary (60) in Section 5.6. It is not hard to see that the law of total probability (25) implies (26), so it is (really) Facts (20), (22), and (25) that are necessary and sufficient for $F : \mathcal{L} \rightarrow [0, 1]$ to represent a probability distribution. We summarize with the following "representation theorem."

(28) FACT: A function $F : \mathcal{L} \rightarrow [0, 1]$ represents a probability distribution if and only if for all $\varphi, \psi \in \mathcal{L}$,

  (a) if $\models \varphi$ then $F(\varphi) = 1$.

  (b) if $\models \varphi \leftrightarrow \psi$ then $F(\varphi) = F(\psi)$.

  (c) $F(\varphi \wedge \psi) + F(\varphi \wedge \neg\psi) = F(\varphi)$.

(29) EXERCISE: Let formulas $\varphi, \psi \in \mathcal{L}$ be given. Show that $\varphi \models \psi$ if and only if for all probability distributions ***Pr***, ***Pr***$(\varphi \wedge \neg\psi) = 0$.

9.3.6 Conditional probability

In considering conditional probability in the context of $\mathcal{L}$ we must beware of a collision in terminology. Our sentential language $\mathcal{L}$ contains conditionals

[16] For discussion, see Fact (54) in Section 5.5.

like $p \rightarrow q$, but we'll see in the next chapter that they do not code conditional events. Rather, conditional events are *pairs* of events, as noted in Section 9.2.6. Thus, in the context of $\mathcal{L}$, conditional events are pairs of subsets of TrAs. When we refer to such events using formulas $\mathcal{L}$, conditional events therefore become pairs of formulas. As previously, we use the symbol $\mid$ as an elongated comma to separate the formulas in a pair. Thus, for $\varphi, \psi \in \mathcal{L}$, the conditional event that φ is true assuming that ψ is true is denoted $\varphi \mid \psi$.

Suppose now that we are given a probability distribution over TrAs. From Definition (17) we know how to apply *Pr* to formulas of $\mathcal{L}$. Now we ask how *Pr* is to be applied to conditional events $\varphi \mid \psi$. It's done by applying Definition (8) in Section 9.3.3 to events in the logical context, as follows.

(30) DEFINITION: Suppose that *Pr* is a probability distribution over TrAs, extended via Definition (17) to $\mathcal{L}$. We extend *Pr* to all pairs of formulas $\varphi \mid \psi$ such that $Pr(\psi) > 0$. For any such pair $\varphi \mid \psi$, we define:

$$Pr(\varphi \mid \psi) = \frac{Pr([\varphi] \cap [\psi])}{Pr([\psi])}.$$

If $Pr(\psi) = 0$ then $Pr(\varphi \mid \psi)$ is not defined.

Unpacking the definition, we see that

$$Pr(\varphi \mid \psi) = \frac{Pr([\varphi] \cap [\psi])}{Pr([\psi])} = \frac{Pr([\varphi \wedge \psi])}{Pr([\psi])} = \frac{Pr(\varphi \wedge \psi)}{Pr(\psi)}.$$

Consequently, (30) implies the following fact, which can be taken as an alternative definition of conditional probability for $\mathcal{L}$.

(31) FACT: Suppose that *Pr* is a probability distribution over $\mathcal{L}$. Then for all pairs of formulas $\varphi \mid \psi$ such that $Pr(\psi) > 0$,

$$Pr(\varphi \mid \psi) = \frac{Pr(\varphi \wedge \psi)}{Pr(\psi)}.$$

To illustrate, suppose the *Pr* is given by (15)*i*. Then

$$\boldsymbol{Pr}(\neg q \mid p \wedge \neg r) = \frac{\boldsymbol{Pr}(\neg q \wedge (p \wedge \neg r))}{\boldsymbol{Pr}(p \wedge \neg r)} = \frac{\boldsymbol{Pr}(\{d\})}{\boldsymbol{Pr}(\{b, d\})} = \frac{.05}{.1 + .05} = \frac{1}{3}.$$

Fact (31) immediately yields another one, quite handy.

(32) FACT: Let distribution *Pr* and $\varphi, \psi \in \mathcal{L}$ be such that $\boldsymbol{Pr}(\psi) > 0$. Then $\boldsymbol{Pr}(\varphi \wedge \psi) = \boldsymbol{Pr}(\psi) \times \boldsymbol{Pr}(\varphi \mid \psi)$.

To illustrate, we follow up the last example:

$$\boldsymbol{Pr}(\neg q \wedge (p \wedge \neg r)) = \boldsymbol{Pr}(\neg q) \times \boldsymbol{Pr}(\neg q \mid p \wedge \neg q) =$$

$$\boldsymbol{Pr}(\{b, d, f, h\}) \times \frac{1}{3} = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6}.$$

The handiest of all facts about conditional probability is *Bayes' Theorem*, stated as follows.

(33) THEOREM: (Bayes' Theorem) Suppose that *Pr* is a probability distribution over $\mathcal{L}$. Then for all pairs of formulas $\varphi \mid \psi$ such that $\boldsymbol{Pr}(\psi) > 0$,

$$\boldsymbol{Pr}(\varphi \mid \psi) = \frac{\boldsymbol{Pr}(\psi \mid \varphi) \times \boldsymbol{Pr}(\varphi)}{\boldsymbol{Pr}(\psi)}.$$

The theorem has many uses (see, e.g., Pearl [79]). Its proof is simple. From Fact (32), $\boldsymbol{Pr}(\varphi \wedge \psi) = \boldsymbol{Pr}(\psi \wedge \varphi) = \boldsymbol{Pr}(\varphi) \times \boldsymbol{Pr}(\psi \mid \varphi) = \boldsymbol{Pr}(\psi \mid \varphi) \times \boldsymbol{Pr}(\varphi)$. So:

$$\boldsymbol{Pr}(\varphi \mid \psi) = \frac{\boldsymbol{Pr}(\varphi \wedge \psi)}{\boldsymbol{Pr}(\psi)} = \frac{\boldsymbol{Pr}(\psi \mid \varphi) \times \boldsymbol{Pr}(\varphi)}{\boldsymbol{Pr}(\psi)}.$$

Sticking with our example, we calculate:

$$\boldsymbol{Pr}((p \wedge \neg r) \mid \neg q) = .1, \quad \boldsymbol{Pr}(\neg q) = .5, \quad \boldsymbol{Pr}(p \wedge \neg r) = .15.$$

Hence by Theorem (33),

$$\boldsymbol{Pr}(\neg q \mid p \wedge \neg r) = \frac{\boldsymbol{Pr}(p \wedge \neg r \mid \neg q) \times \boldsymbol{Pr}(\neg q)}{\boldsymbol{Pr}(p \wedge \neg r)} = \frac{.1 \times .5}{.15} = \frac{1}{3},$$

which is the same value we obtained earlier when we calculated $\boldsymbol{Pr}(\neg q \mid p \wedge \neg r)$ directly from Fact (31).

By the law of total probability (25), $\boldsymbol{Pr}(\psi) = \boldsymbol{Pr}(\varphi \wedge \psi) + \boldsymbol{Pr}(\neg\varphi \wedge \psi)$. By (32), $\boldsymbol{Pr}(\varphi \wedge \psi) = \boldsymbol{Pr}(\varphi) \times \boldsymbol{Pr}(\psi \mid \varphi)$, and $\boldsymbol{Pr}(\neg\varphi \wedge \psi) = \boldsymbol{Pr}(\neg\varphi) \times \boldsymbol{Pr}(\psi \mid \neg\varphi)$. Putting these facts together with (33) yields another version of the theorem, often seen:

(34) THEOREM: (Bayes' Theorem, expanded version) Suppose that *Pr* is a probability distribution over $\mathcal{L}$. Then for all pairs of formulas $\varphi \mid \psi$ such that $\boldsymbol{Pr}(\psi) > 0$,

$$\boldsymbol{Pr}(\varphi \mid \psi) = \frac{\boldsymbol{Pr}(\psi \mid \varphi) \times \boldsymbol{Pr}(\varphi)}{(\boldsymbol{Pr}(\psi \mid \varphi) \times \boldsymbol{Pr}(\varphi)) + (\boldsymbol{Pr}(\psi \mid \neg\varphi) \times \boldsymbol{Pr}(\neg\varphi))}.$$

9.3.7 Coherence

Consider a pair (φ, x) consisting of a formula φ, and a number x. In this section (which can be skipped) we'll consider such a pair to be the affirmation that the probability of φ is x. Similarly, given a triple (χ, ψ, y), with $\chi, \psi \in \mathcal{L}$ and y a number, we interpret the triple as the affirmation that the probability of χ assuming ψ is y. Call any such pair or triple a *probability claim*.

(35) DEFINITION: Let

$$C = \{(\varphi_1, x_1), \cdots (\varphi_n, x_n), (\chi_1, \psi_1, y_1), \cdots (\chi_m, \psi_m, y_m)\}$$

be a collection of probability claims. Then C is called *coherent* just in case there is a probability distribution *Pr* for $\mathcal{L}$ such that

$$\boldsymbol{Pr}(\varphi_1) = x_1, \cdots \boldsymbol{Pr}(\varphi_n) = x_n$$

and

$$\textbf{\textit{Pr}}(\chi_1 \,|\, \psi_1) = y_1, \cdots \textbf{\textit{Pr}}(\chi_m \,|\, \psi_m) = y_m.$$

If there is no such probability distribution then C is called *incoherent*.

Coherence requires there to be at least one probability distribution ***Pr*** that returns the right numbers on *all* the pairs and triples (it's not good enough that different distributions work for different pairs or triples).

(36) EXAMPLE: Consider the following three sets of probability claims.

(a) $\{(p, .3), (\neg q \vee p, .4)\}$
(b) $\{(p, .3), (\neg q \vee p, .3)\}$
(c) $\{(p, .3), (\neg q \vee p, .2)\}$

You should be able to see that only the first two are coherent. Set (c) is incoherent because $[p] \subseteq [\neg q \vee p]$, hence, greater probability cannot be assigned to p compared to $\neg q \vee p$.

(37) EXAMPLE: For another illustration, consider the following three sets of probability claims.

(a) $\{(p, .8), (q \wedge p, .9)\}$
(b) $\{(p, .8), (q \wedge p, .8)\}$
(c) $\{(p, .8), (q \wedge p, .7)\}$

In this case, it is the first set of claims that is incoherent; the other two are coherent.

(38) EXAMPLE: Finally, consider the following set of three probability claims.

$$\{(p, .8), (q, p, .5), (q \wedge p, .3)\}$$

By Definition (8), this set is incoherent since for any distribution ***Pr***,

$$\textbf{\textit{Pr}}(q \,|\, p) = \frac{\textbf{\textit{Pr}}(q \wedge p)}{\textbf{\textit{Pr}}(p)} = \frac{.3}{.8} \neq .5.$$

As our terminology suggests, it is a sin to advance an incoherent set of probability claims. For one thing, it's a misuse of the technical term "probability," since the numbers don't conform to any (genuine) probability distribution. Another reason to avoid incoherent probability claims is that you might be challenged to accept bets corresponding to them. Bad things can happen to someone who accepts bets that seem fair according to incoherent probabilities. We won't tell that story here; see Resnik [83, Ch. 3-3c] instead.

9.4 Independence

Let us touch briefly on the topic of independence.[17]

(39) DEFINITION: Let $\varphi, \psi \in \mathcal{L}$ and probability distribution *Pr* be given. We say that φ and ψ are *independent with respect to Pr* just in case $\boldsymbol{Pr}(\varphi \mid \psi) = \boldsymbol{Pr}(\varphi)$.

Note that formulas are independent only with respect to a particular distribution. Often it is clear which distribution is intended, and reference to it is left implicit. We leave the proof of the following fact to you.

(40) FACT: Let $\varphi, \psi \in \mathcal{L}$ and probability distribution *Pr* be given.

(a) φ and ψ are independent with respect to *Pr* if and only if ψ and φ are independent with respect to *Pr*.[18]

(b) φ and ψ are independent with respect to *Pr* if and only if $\boldsymbol{Pr}(\psi \wedge \varphi) = \boldsymbol{Pr}(\psi) \times \boldsymbol{Pr}(\varphi)$.

Finally, we observe that independence is not a transitive relation. That is, if p and q are independent with respect to *Pr*, and q and r are independent with respect to the same distribution *Pr*, it does not follow that p and r are

[17] For much more, see [15, 80].

[18] That is, the relation *being independent of* is symmetric. This fact does not follow from the mere choice of terminology. It must be proved from Definition (39).

independent with respect to ***Pr***. For an example of nontransitivity, consider two fair coins. The first has the letters P and R inscribed on one side, blank on the other. The second has the letter Q inscribed on one side, blank on the other. The coins are tossed separately, and we examine the revealed faces for letters. Let p, q, r be the assertion that P, Q, and R appear, respectively. For the distribution ***Pr*** we've described, it is clear that $\boldsymbol{Pr}(p \mid q) = \boldsymbol{Pr}(p) = \frac{1}{2}$, $\boldsymbol{Pr}(q \mid r) = \boldsymbol{Pr}(q) = \frac{1}{2}$, but $\boldsymbol{Pr}(p \mid r) = 1 \neq \frac{1}{2} = \boldsymbol{Pr}(p)$. So, p and q are independent, as are q and r. But p and r are dependent.

That's all you need to know about probability to resume consideration of conditionals (actually, it was a bit more than you need). You may be tired after this long excursion through *inductive logic*. To get ready for a triumphal return to deductive logic, take our advice: rest up, have a good (but low-calorie) meal, and think of nothing else but truth-assignments for the next 24 hours.

Chapter 10

A theory of indicative conditionals

We're back! You seem to be back too. Thanks for joining us in Chapter 10.

Let's see ... Where were we? Oh yes. In Chapter 8 we reached the nadir of our fortune, having apparently demonstrated that → perspicuously represents *if–then–* in English, and also that → fails to perspicuously represent *if–then–* in English. The present chapter is devoted to exploring one potential solution to this mystery. In fact, many different ideas have been advanced by philosophers and linguists to explain the meaning of indicative conditionals. The approach we favor is similar to that of Bruno de Finetti [23], rediscovered (and more fully developed) by Michael McDermott [71]. Other accounts along similar lines include Adams (1998) and Bennett (2003).[1] Our theory differs from theirs in various ways, however. So the reader should attribute anything that seems confused or confusing to the current authors, not to anyone else. We also plunder ideas from the framework known as *supervaluationism*, explained in Beall and van Fraassen [10, Ch. 9].

One motivation for our approach is its consonance with the apparatus of Sentential Logic constructed in Chapters 4 and 5. For an important perspective alternative to the one discussed here, see Lycan [69]; the same work provides illuminating discussion of yet other theories. It's essential to keep such alternatives in mind since the theory to be explored in this chapter is not entirely satisfactory (but we're getting ahead of the story).

Before explaining the central idea of our theory, let us first consider a tempting theory of *if–then–* that we intend to reject, or rather, transform into something more palatable. The digression will be long, however. If you're impatient to get to the heart of the matter, skip to Section 10.2.

10.1 Conditionals deprived of truth value

10.1.1 One way to resolve the paradox

Maybe an indicative conditional like

[1] Additional historical antecedents to the theory are described in [71].

(1) If Schwarzenegger is reelected governor in 2007 then he'll be elected president in 2008.

doesn't have a truth-value. Following Lycan [69], let us call this thesis *NTV* ("no truth value"). You can believe NTV without taking (1) to be meaningless. Simply, the meaning does not make the sentence either true or false. Of course, the left hand side and the right hand side of (1) have truth values. Moreover, (1) seems to relate its two sides in a conditional way, but without giving the whole sentence a truth-value (according to NTV). If you think that genuine propositions must be either true or false, you can express this idea by saying that (1) does not express a conditional proposition but rather expresses a proposition conditionally.[2] The proposition expressed conditionally is that Schwarzenegger will be elected president in 2008; the condition that must be met for this proposition to be expressed is that he is reelected governor in 2007. Since we're not sure whether propositions must, by definition, have truth-values, we'll just interpret NTV as the thesis that (1) is without one.[3]

According to NTV (or at least, the version of the thesis that we are considering), the truth-value-less character of (1) is not due to its reference to future events. The following sentence about the distant past would also be without truth value.

(2) If Mars had liquid surface water in its first billion years then life once flourished there.

Again, the idea is that (2) expresses the claim (either true or false) that life once flourished on Mars, but it expresses this claim just in case it is true that Mars had liquid water in its first billion years. The sentence as whole, however, is neither true nor false.[4]

[2]This turn of phrase is due to W. V. O. Quine [82, p. 21].

[3]NTV is developed in [5, 27, 108]. What follows exposes just a fraction of the ideas advanced to support the thesis. For the balance, we invite you to consult the literature on interpretations of the indicative conditional, starting with overviews like [108, 69, 78].

[4]For more nuanced views of the interaction of time and conditionals, see Jackson [51] and Dancygier [22].

In one stroke, NTV dissolves the contradictory results of Chapter 8. Those results hinged on comparing validity in $\mathcal{L}$ with secure inference (symbolized by $\Rightarrow$). You'll recall from Section 8.4.1 that we write $\{A_1 \cdots A_n\} \Rightarrow B$ just in case it is not possible for all of $A_1 \cdots A_n$ to be true yet C be false. This definition seems not to be adapted to our concerns about secure inference when it comes to statements without truth-values. For example, let E be "Ducks dance." Then we have $(1) \Rightarrow E$ because indicative conditionals have no truth value (according to NTV), hence (1) *can't* be true, so it can't be true while Ducks fail to dance. Thus, the definition of $\Rightarrow$ rules the argument from (1) to E to be secure, which is preposterous. So, if NTV is right, we cannot trust reasoning that blends indicative conditionals and $\Rightarrow$. This is precisely the defect (according to NTV) that undermines the entire discussion of Chapter 8. For example, to argue that *if–then–* is not transitive, we offered the Queen Elizabeth example (49) in Section 8.5.2. Both the conclusion and the premises were conditionals so all reference to secure inference was pointless. The security in question is supposed to ensure that true premises lead to true conclusions. But in the example, none of the statements are either true or false!

The Queen Elizabeth example was used to demonstrate that *if–then–* cannot be modeled by $\rightarrow$. For the other side of the paradox, we relied on supposed facts about $\Rightarrow$ to show that *if–then–* can be so modeled after all. For example, in Section 8.4.1 we used the following.

> FIRST CONDITIONAL PRINCIPLE FOR ENGLISH: For every pair A, B of sentences, *if-A-then-B* $\Rightarrow$ *not-*(A*-and-not-*B)

Once again, this principle reduces to an unintended triviality should it be the case (as urged by NTV) that *if-A-then-B* has no truth value. So use of the principle undermines the demonstration we presented in favor of $\rightarrow$ as a model of *if–then–*. The same illicit mixture of conditionals and $\Rightarrow$ infects all the arguments used to support the two sides of the paradox. So those arguments can be discounted. The paradox is thereby dissolved by identifying an untenable assumption common to the opposing arguments. The common assumption (false, according to NTV) is that English conditionals have a truth value.

But perhaps NTV seems incredible to you. Could a declarative sentence like (1) really fail to be either true or false, even though its left hand side and right hand side indisputably *do* have truth values? How come such a thing doesn't happen to sentences with other connectives like "and" and "or"? For example, the first two sentences in the following list certainly seem to be either true or false; is it credible that the third is so radically different?

(3) (a) Chipmunks live on Venus and chipmunks don't mind heat.
    (b) Chipmunks live on Venus or chipmunks don't mind heat.
    (c) If chipmunks live on Venus then chipmunks don't mind heat.

But the superficial grammatical similarity of (3)c to (3)a,b may be misleading. In some ways, the *if–then–* construction in English is unlike constructions involving *and* and *or*. Notice, for example, that (3)a,b can be reduced as shown in (4) whereas this is not possible for (3)c.

(4) (a) Chipmunks live on Venus and don't mind heat.
    (b) Chipmunks live on Venus or don't mind heat.
    (c) *If chipmunks live on Venus then don't mind heat.

Putting the * in front of (4)c signifies its ungrammaticality in English, which contrasts with the grammaticality of (4)a,b. The special grammar of "if" also shows up in queries. Thus, the following transformations of (3)a,b are ungrammatical.

(5) (a) *What lives on Venus and chipmunks don't mind heat?
    (b) *What lives on Venus or chipmunks don't mind heat?

In contrast, the same kind of transformation successfully converts (6)a below into the query (6)b.

(6) (a) Chipmunks don't mind heat if chipmunks live on Venus.
    (b) What doesn't mind heat if chipmunks live on Venus?

On the other hand, a slightly different kind of query is allowed for (3)a,b but not (3)c. Witness:

(7) (a) What lives on Venus and doesn't mind heat?

(b) What lives on Venus or doesn't mind heat?

(c) *What if lives on Venus then doesn't mind heat? [Also: *If what lives on Venus then doesn't mind heat?]

Yet other grammatical distinctions between conditionals and related constructions are discussed in Lycan [69]. So perhaps conditionals are grammatically peculiar, which might make NTV seem more plausible.

10.1.2 Another reason to doubt that conditionals have truth values

To provide more direct evidence in favor of NTV, suppose it to be wrong. That is, suppose that conditionals like (1) have truth values after all. Then they must have probability. For, any sentence that can be true has some chance of actually being true. In the "subjectivist" framework explained in Section 9.2.4, there must therefore be some sensible estimate of the chance of, for example, (1). Let's write this thought down.

(8) If NTV is false then *Pr*(*if*-A-*then*-B) is well defined for any statements A, B with determinate truth values [with the proviso that *Pr*(A) is positive].

Now, in Section 9.3.6 we considered "conditional probability," and warned about a collision of terminology. Conditional statements are one thing, conditional probability another. Yet we describe the number *Pr*($B \mid A$) as "the probability of B assuming A." The latter expression doesn't seem far from "the probability of B if A," hence, it seems close to "the probability of *if*-A-*then*-B." Consider an example. Suppose that we are about to throw a fair, six-sided die. What probability feels right for the following conditional?

(9) If the die shows an even number then the die shows 6.

It sure seems that the probability of (9) is $1/3$ since 6 is one of the three (equally likely) ways an even number could turn up. And $1/3$ is also the conditional probability of 6 given even, for:

$$\mathit{Pr}(6 \mid \text{even}) = \frac{\mathit{Pr}(6 \text{ and even})}{\mathit{Pr}(\text{even})} = \frac{\mathit{Pr}(6)}{\mathit{Pr}(\text{even})} = \frac{1/6}{1/2} = \frac{1}{3}.$$

We are led in this way to the hypothesis that $\mathit{Pr}(B \mid A) = \mathit{Pr}(\textit{if-}A\textit{-then-}B)$, in brief, that conditional probability is the probability of a conditional. (This idea was crisply formulated in Stalnaker [94].) In view of (8), we now have:

(10) If NTV is false then for any statements A, B with determinate truth values, $\mathit{Pr}(\textit{if-}A\textit{-then-}B) = \mathit{Pr}(B \mid A)$, where *Pr* is whatever distribution happens to govern the probability of statements in English [and provided that $\mathit{Pr}(A) > 0$].

But the right hand side of (10) should look suspicious to you in light of Theorem (10) of Section 9.2.7. There it was proved that (roughly speaking) no function maps pairs of events into single events whose probabilities correspond to the conditional probabilities of the pairs. It seems that (10) is likewise flirting with the impossible, if NTV is false. The earlier theorem involved the probability of events defined in a sample space rather than the probabilities of statements in a language like English. So let us revisit the theorem in the present setting.

To see more clearly what is at issue, let us temporarily retreat back to consideration of $\mathcal{L}$ instead of English. We ask whether $p \to q$ has the following property.

(11) For all distributions *Pr* over $\mathcal{L}$, $\mathit{Pr}(p \to q) = \mathit{Pr}(q \mid p)$.

It is tempting to believe (11) because $p \to q$ is called a "conditional" and pronounced "if p then q." But to see that (11) is wrong it suffices to calculate $\mathit{Pr}(q \mid p)$ and $\mathit{Pr}(p \to q)$ according to the following distribution [also seen as (15)i of Section 9.3.2].

	p	q	r	*prob*
(a)	t	t	t	.1
(b)	t	t	f	.1
(c)	t	f	t	.2
(d)	t	f	f	.1
(e)	f	t	t	.15
(f)	f	t	f	.1
(g)	f	f	t	.2
(h)	f	f	f	.05

In this case, we have:

$$\mathbf{Pr}(p \to q) = \mathbf{Pr}(\{a, b, e, f, g, h\}) = .1 + .1 + .15 + .1 + .2 + .05 = .7.$$

$$\mathbf{Pr}(q \mid p) = \frac{\mathbf{Pr}(p \wedge q)}{\mathbf{Pr}(p)} = \frac{\mathbf{Pr}(a, b)}{\mathbf{Pr}(a, b, c, d)} = \frac{.1 + .1}{.1 + .1 + .2 + .1} = \frac{.2}{.5} = .4.$$

So (11) is wrong. In fact, (11) is even "qualitatively" wrong. You know that $\models (p \to q) \leftrightarrow (\neg q \to \neg p)$. So by Fact (22) in Section 9.3.4, for all distributions $\mathbf{Pr}$, $\mathbf{Pr}(p \to q) = \mathbf{Pr}(\neg q \to \neg p)$. Hence, if (11) were true, it would be the case that $\mathbf{Pr}(q \mid p) = \mathbf{Pr}(\neg p \mid \neg q)$ for all distributions $\mathbf{Pr}$. But this is not the case. Suppose you throw a fair coin twice. Let q be the claim that at least one toss comes up heads. Let p be the claim that at least one toss comes up tails. Then it is easy to calculate that $\mathbf{Pr}(q \mid p) = \frac{2}{3}$ whereas $\mathbf{Pr}(\neg p \mid \neg q) = 0$.[5]

Although (11) is wrong, we can still wonder whether there is some formula alternative to $p \to q$ that does the trick. Perhaps for all distributions $\mathbf{Pr}$, $\mathbf{Pr}(q \mid p) = \mathbf{Pr}(p \vee \neg q)$, or $\mathbf{Pr}(q \mid p) = \mathbf{Pr}(\neg p \wedge (q \to p))$, for example. Or perhaps some variable other than p and q needs to enter the picture. Thus, we must consider the hypothesis that $\mathbf{Pr}(q \mid p) = \mathbf{Pr}(p \to (z \vee q))$, or $\mathbf{Pr}(q \mid p) =$

[5] A more intuitive example is due to Cohen [20, p. 21]. The conditional probability that a randomly chosen person lives in Oxford given that he lives in England is quite low (because there are so many other places to live in England). But the conditional probability that a randomly chosen person does not live in England given that they do not live in Oxford is quite high (because there are so many other countries to live in). Further distinctions among constructions called "conditionals" are discussed in [20, §3].

$Pr((p \leftrightarrow z) \vee (w \rightarrow q))$, or that $Pr(q \mid p)$ equals some other weird formula. Even this is not general enough. It is possible that our language $\mathcal{L}$ is too impoverished to express conditional probability, but that it would be possible with the introduction of some new connectives (to supplement $\neg$, $\wedge$, $\vee$, $\rightarrow$, and $\leftrightarrow$).

To address the issue generally, we'll show that no formula in *any* reasonable language can play the role that $p \rightarrow q$ in (11) was supposed to play for $\mathcal{L}$. So in particular, *if-p-then-q* doesn't play this role for English. From (10), we can therefore conclude that NTV is true (since assuming its falsity leads to falsity). Such is the form of our second piece of evidence in favor of NTV. The discussion that follows is based on [67, 13].

Let us first be more specific about the language in which probabilities are being expressed. Of course, English (or some other natural language) is what interests us. But it will be more convenient to consider instead an arbitrary extension of $\mathcal{L}$. Specifically, let $\mathcal{L}^*$ be a language that includes $\mathcal{L}$ as a subset (that is, every formula of $\mathcal{L}$ is also a formula of $\mathcal{L}^*$). We need to make some further assumptions about $\mathcal{L}^*$ but when we're finished it should be clear that $\mathcal{L}^*$ could be chosen to be a healthy fraction of English itself. Hence, showing that $\mathcal{L}^*$ doesn't have the resources to express conditional probabilities with a single formula will be enough to persuade us of the same thing about English.

In particular, we assume that $\mathcal{L}$ has at least two variables, p, q, so $\mathcal{L}^*$ (which extends $\mathcal{L}$) also includes these two variables. It is also assumed that $\mathcal{L}^*$ comes equipped with a relation of logical implication, which we'll call $\models$, just like for $\mathcal{L}$. That is, we only consider extending $\mathcal{L}$ to a language for which it is clear which formulas guarantee the truth of which other formulas. We also assume that probabilities can be sensibly distributed to the formulas of $\mathcal{L}^*$. Specifically, we shall consider a function $Pr : \mathcal{L}^* \rightarrow [0, 1]$ to be a genuine probability distribution only if the restriction of Pr to $\mathcal{L}$ is a probability distribution in the original sense of Section 9.3.2. In other words, given a probability distribution for the new-fangled language $\mathcal{L}^*$, we must be able to recover an old-fashion probability distribution by ignoring all the formulas in $\mathcal{L}^* - \mathcal{L}$. This is not quite all that we need to assume about probability distributions over $\mathcal{L}^*$. We also require:

(12) ASSUMPTIONS ABOUT PROBABILITY DISTRIBUTIONS OVER $\mathcal{L}^*$:

(a) For all $\theta, \psi \in \mathcal{L}^*$, if $\psi \models \theta$ then for all probability distributions ***Pr*** over $\mathcal{L}^*$, ***Pr***$(\psi) \leq$ ***Pr***(θ) [as in Fact (21) of Section 9.3.4].

(b) For all $\varphi \in \mathcal{L}^*$, if $\varphi \not\models p$ then there is some probability distribution ***Pr*** over $\mathcal{L}^*$ such that:

   i. ***Pr***$(\varphi \wedge \neg p) > 0$

   ii. ***Pr***$(p \wedge \neg q) > 0$.

   iii. ***Pr***$(p \wedge q) = 0$.

These are not particularly restrictive assumptions. If $\mathcal{L}^* = \mathcal{L}$, they are clearly met. Assumption (12)a is natural for any reasonable language $\mathcal{L}^*$. It will be clearer to you that assumption (12)b is also reasonable if you observe that $\neg p$, $p \wedge \neg q$ and $p \wedge q$ are pairwise inconsistent (each implies the negation of the others). If we assume that $\varphi \not\models p$ then (12)bi must be possible for some distribution ***Pr***, which can easily be adapted to satisfy (12)bii,biii in view of the incompatibility of $\neg p$, $p \wedge \neg q$ and $p \wedge q$.

To proceed, let us say that a formula $\varphi \in \mathcal{L}^*$ *expresses conditional probability* just in case for all probability distributions ***Pr*** for $\mathcal{L}^*$, ***Pr***$(q \,|\, p) =$ ***Pr***(φ). Our question is: Does any formula of $\mathcal{L}^*$ express conditional probability? Notice that we are focussing attention on just the conditional $(q \,|\, p)$; only p and q are involved. It is clear, however, that a negative answer in this simple case shows that no formula expresses conditional probability more generally. Under the assumptions (12), we now demonstrate:

(13) FACT: No formula of $\mathcal{L}^*$ expresses conditional probability.[6]

Proof:[7] Choose any formula $\varphi \in \mathcal{L}^*$. To demonstrate (13) it must be shown that:

(14) For some probability distribution ***Pr***, ***Pr***$(\varphi) \neq$ ***Pr***$(q \,|\, p)$.

[6]By taking $\mathcal{L}^*$ to be the *null extension*, that is, $\mathcal{L}$ itself, the fact also shows that no formula of $\mathcal{L}$ expresses conditional probability. The same fact is an easy corollary of Theorem (10) (due to David Lewis) in Section 9.2.7.

[7]Once again, our proof is an adaptation of Bradley [13].

We distinguish two cases: $\varphi \models p$, and $\varphi \not\models p$. We'll show that in both cases there is a distribution *Pr* that satisfies (14). Suppose first that $\varphi \models p$. Choose any probability distribution *Pr* such that $0 < \mathbf{Pr}(p \wedge q) = \mathbf{Pr}(p) < 1$. Of course, such distributions exist since *Pr* is an old-fashioned distribution over $\mathcal{L}$ [we assumed this just above (12)]. Then

$$\mathbf{Pr}(q \mid p) = \frac{\mathbf{Pr}(p \wedge q)}{\mathbf{Pr}(p)} = 1.$$

But $\mathbf{Pr}(\varphi) \neq 1$ since otherwise by (12)a, $\mathbf{Pr}(p) = 1$ because $\varphi \models p$ [and $\mathbf{Pr}(p) = 1$ would contradict our choice of *Pr*]. So $\mathbf{Pr}(\varphi) \neq \mathbf{Pr}(q \mid p)$, satisfying (14) as promised.

Now suppose that $\varphi \not\models p$. Then by (12) we may choose a distribution *Pr* that satisfies (12)b. From (12)bi and (12)a, and the fact that $\varphi \wedge \neg p \models \varphi$, $\mathbf{Pr}(\varphi) > 0$. By the same reasoning, from (12)bii it follows that $\mathbf{Pr}(p) > 0$. So by (12)biii we obtain:

$$\frac{\mathbf{Pr}(p \wedge q)}{\mathbf{Pr}(p)} = \mathbf{Pr}(q \mid p) = 0.$$

Hence, in this case too, $\mathbf{Pr}(\varphi) \neq \mathbf{Pr}(q \mid p)$. ∎

Let us recall the significance of Fact (13) in the larger discussion of indicative conditionals. We saw in Section 10.1.1 above that one way to resolve the conflicting arguments in Chapter 8 is to assume that indicative conditionals like (1) have no truth value. This thesis was called NTV. To bolster NTV, we tried to convince you that if conditionals *do* have truth values then their probabilities are the conditional probabilities of their right hand side given their left hand side. Then we presented a celebrated argument that this is impossible. We therefore concluded that indicative conditionals don't have truth values, agreeing thereby with NTV.

But now we'll provide good reasons for nonetheless doubting NTV!

10.1.3 Against NTV

Look again at (10), the pivot of our second reason for believing NTV. Didn't you think it was true? Or maybe we didn't convince you, and you thought it was false. Or maybe you couldn't decide whether it was true or false. In any case, we suspect that it never crossed your mind that (10) was neither true nor false. So you don't really believe NTV, which claims that conditionals like (10) lack truth value!

To press the point, consider the following case. You finally think of a joke and offer to tell it to us provided we promise to laugh. We might assert any of:

(15) (a) If you tell your joke then we'll laugh.

(b) We'll laugh when you tell your joke.

(c) We'll laugh in the event that you tell your joke.

(d) We'll laugh at the telling of your joke.

(e) We'll laugh should you tell your joke.

So you tell your joke and we don't move a muscle. Aren't you justified in crying *Liars!* no matter which assertion from (15) we happened to make? Surely we could not (reasonably) defend ourselves in the specific case (15)a by denying it a truth value. ("We didn't say anything false, you see, since indicative conditionals are neither true nor false.") If (15)b-e have truth values then so does (15)a. Don't you think that's true (and that it's a conditional)? And denying truth values to (15)b-e seems like a desperate defense of NTV.[8] Also, in a true/false math test, would you dare mark the following assertion as *neither*?

(16) If Mercury is almost a perfect sphere then its circumference exceeds its diameter.

An advocate of NTV wishing to pass Math 101 might wish to consider (16) as special for some reason, perhaps because numbers are involved. But it is hard to see why it should be treated differently from the equally true:

[8] We have adapted this argument from [69, Ch. 4], and likewise for the arguments appearing in the rest of Section 10.1.3.

> If Mercury is almost a perfect sphere then it was molten at some time.

For a related example, consider the conditional:

(17) If the Statue of Liberty is made of bronze then it conducts electricity.

It seems difficult to deny the truth of (17) since it is an indisputable consequence of a sentence that is indisputably true, namely:

> Everything bronze conducts electricity.

Thus, to maintain NTV it is necessary to deny that the consequence of a true sentence must be true.

For the reasons just rehearsed, let us abandon NTV and grant that conditionals may often have truth values. The theory we'll now develop, however, cedes a kernel of truth to NTV since it countenances truth "gaps" for conditionals in certain cases. We'll also find a kernel of truth in the claim that $\boldsymbol{Pr}(\textit{if-}A\textit{-then-}B) = \boldsymbol{Pr}(B \mid A)$.

10.2 A theory based on truth gaps

The theory will be described in the present section, then its consequences discussed in Section 10.4. The general idea is to amend Sentential Logic so that $\rightarrow$ in the revised system successfully represents *if–then–* in English.

10.2.1 Truth conditions and falsity conditions

Let's call the set of truth-assignments that make a given formula φ true, the *truth conditions* of φ. To illustrate with a familiar case, here again is a list of the eight truth-assignments that issue from three variables. (We first met this list in Section 4.2.1.)

(18)

	p	q	r
(a)	T	T	T
(b)	T	T	F
(c)	T	F	T
(d)	T	F	F
(e)	F	T	T
(f)	F	T	F
(g)	F	F	T
(h)	F	F	F

Thus, the truth conditions of $(p \wedge q) \vee r$ is the set $\{$(a), (b), (c), (e), (g)$\}$.

Let's call the set of truth-assignments that make a given formula φ false, the *falsity conditions* of φ. Thus, the falsity conditions of $(p \wedge q) \vee r$ is the set $\{$(d), (f), (h)$\}$. It is no accident that the falsity conditions of φ are complementary to its truth conditions. The definition of a truth-assignment satisfying a formula, given in Section 4.2.2, was designed to ensure such an outcome. For every formula φ, and every truth-assignment α, either $\alpha \models \varphi$ (in which case α is one of the truth conditions of φ) or $\alpha \models \neg\varphi$ (in which case α is one of the falsity conditions of φ). Let us explore the consequences of changing this assumption in the case of $\rightarrow$. Specifically, we assume that a truth-assignment α assigns no truth value to $\chi \rightarrow \psi$ if α assigns F to χ. That is, if $\alpha(\chi) = \mathsf{F}$ then $\alpha(\chi \rightarrow \psi)$ is *undefined*.[9] The latter stipulation concerns how a truth-assignment is extended from the variables of $\mathcal{L}$ to non-atomic formulas (notably, to conditionals). The "core" concept of a truth-assignment is not altered. It is still a (total) mapping of each sentential variable into $\{\mathsf{T}, \mathsf{F}\}$. A truth-assignment is never undefined on a variable (but it *is* sometimes undefined on nonatomic formulas).

To make these ideas precise, let us reformulate the semantical concepts introduced in Section 4.2.2 by modifying Definition (6). The original definition shows how a given truth-assignment can be extended from the variables to all of $\mathcal{L}$. The new definition will extend a truth-assignment only to a subset (still infinite) of $\mathcal{L}$; the truth-assignment will be undefined on many formulas. We state the new definition in terms of a given truth-assignment α, rather than

[9] Quine [81, p. 226] acknowledges a truth-value "gap" in English conditionals when their left hand side is false. For more history of the same idea, see [71].

in terms of its extension $\overline{\alpha}$. Recall from Definition (13) of Section 4.2.3 that we allow a truth-assignment to refer to its own extension.

(19) DEFINITION: Suppose that a truth-assignment α and a formula φ are given. φ is either atomic, a negation, a conjunction, a disjunction, a conditional, or a biconditional. We define $\alpha(\varphi)$ in all these cases.

(a) Suppose that φ is the atomic formula v_i. Then $\alpha(\varphi)$ is already defined, and $\alpha(\varphi)$ is either T or F (and not both, obviously).

(b) Suppose that φ is the negation $\neg\psi$. Then $\alpha(\varphi) = \mathsf{T}$ if $\alpha(\psi) = \mathsf{F}$, and $\alpha(\varphi) = \mathsf{F}$ if $\alpha(\psi) = \mathsf{T}$. If $\alpha(\psi)$ is not defined, then neither is $\alpha(\varphi)$. [That is, if $\alpha(\psi)$ is neither T nor F according to α, then $\neg\psi$ is likewise neither T nor F according to α.]

(c) Suppose that φ is the conjunction $\chi \wedge \psi$. Then $\alpha(\varphi) = \mathsf{T}$ just in case $\alpha(\chi) = \mathsf{T}$ and $\alpha(\psi) = \mathsf{T}$. If either $\alpha(\chi) = \mathsf{F}$ or $\alpha(\psi) = \mathsf{F}$, then $\alpha(\varphi) = \mathsf{F}$. In all other cases, $\alpha(\varphi)$ is not defined.

(d) Suppose that φ is the disjunction $\chi \vee \psi$. Then $\alpha(\varphi) = \mathsf{F}$ just in case $\alpha(\chi) = \mathsf{F}$ and $\alpha(\psi) = \mathsf{F}$. If either $\alpha(\chi) = \mathsf{T}$ or $\alpha(\psi = \mathsf{T}$, then $\alpha(\varphi) = \mathsf{T}$. In all other cases, $\alpha(\varphi)$ is not defined.

(e) Suppose that φ is the conditional $\chi \rightarrow \psi$. Then $\alpha(\varphi) = \mathsf{T}$ just in case either (i) $\alpha(\chi) = \mathsf{T}$ and $\alpha(\psi) = \mathsf{T}$, or (ii) $\alpha(\chi)$ is undefined and $\alpha(\psi) = \mathsf{T}$. $\alpha(\varphi) = \mathsf{F}$ just in case either (i) $\alpha(\chi) = \mathsf{T}$ and $\alpha(\psi) = \mathsf{F}$, or (ii) $\alpha(\chi)$ is undefined and $\alpha(\psi) = \mathsf{F}$. In all other cases $\alpha(\varphi)$ is not defined.

(f) Suppose that φ is the biconditional $\chi \leftrightarrow \psi$. Then $\alpha(\varphi) = \mathsf{T}$ just in case $\alpha(\chi) = \mathsf{T}$ and $\alpha(\psi) = \mathsf{T}$. $\alpha(\varphi) = \mathsf{F}$ just in case either (i) $\alpha(\chi) = \mathsf{T}$ and $\alpha(\psi) = \mathsf{F}$, (ii) $\alpha(\chi) = \mathsf{F}$ and $\alpha(\psi) = \mathsf{T}$, (iii) $\alpha(\chi)$ is undefined and $\alpha(\psi) = \mathsf{F}$, or (iv) $\alpha(\chi) = \mathsf{F}$ and $\alpha(\psi)$ is undefined. In all other cases $\alpha(\varphi)$ is not defined.

Definition (19) is easier to remember than it seems, since its clauses follow a pattern that we will shortly explain. First, let us summarize the definition via the following truth tables for $\mathcal{L}$'s five connectives. The symbol U signifies "undefined."

(20) NEW TABLE FOR NEGATION:

$\neg$	ψ
F	T
U	U
T	F

(21) NEW TABLE FOR CONJUNCTION:

χ	$\wedge$	ψ
T	T	T
T	F	F
T	U	U
U	U	T
U	F	F
U	U	U
F	F	T
F	F	F
F	F	U

(22) NEW TABLE FOR DISJUNCTION:

χ	$\vee$	ψ
T	T	T
T	T	F
T	T	U
U	T	T
U	U	F
U	U	U
F	T	T
F	F	F
F	U	U

(23) NEW TABLE FOR CONDITIONALS:

χ	$\rightarrow$	ψ
T	T	T
T	F	F
T	U	U
U	T	T
U	F	F
U	U	U
F	U	T
F	U	F
F	U	U

(24) NEW TABLE FOR BICONDITIONALS:

χ	$\leftrightarrow$	ψ
T	T	T
T	F	F
T	U	U
U	U	T
U	F	F
U	U	U
F	F	T
F	U	F
F	F	U

(25) EXAMPLE: Let α be (c) in Table (18). Then $\alpha(\neg(q \rightarrow p))$ is undefined because $\alpha(q \rightarrow p)$ is undefined [because $\alpha(q) = \mathsf{F}$]. On the other hand, if α is (b) in Table (18). then $\alpha(\neg(q \rightarrow p)) = \mathsf{F}$.

(26) EXAMPLE: Let α be (d) in Table (18). Then $\alpha((p \wedge q) \leftrightarrow (p \rightarrow q)) = \mathsf{U}$ because $\alpha(p \wedge q) = \mathsf{F}$ and $\alpha(p \rightarrow q) = \mathsf{F}$. On the other hand, if α is (b) in Table (18) then $\alpha((p \wedge q) \leftrightarrow (p \rightarrow q)) = \mathsf{T}$ because (b) assigns T to both sides of the biconditional.

10.2.2 Interpreting the new truth tables

Definition (19) envisions only two truth values, T and F, just like the original Definition (6) in Section 4.2.2. In particular, we do not conceive of U ("undefined") as a new, third truth value. Rather, when $\alpha(\varphi)$ is undefined, α assigns nothing at all to φ. To insist on this point, we might have used a blank in place of U, but the tables would be harder to read.

To remember the new truth tables, observe that they are consistent with the truth tables from standard logic in the following sense. Each row in one of the new tables corresponds to some row in the old table for the same connective except that one or more Ts and Fs have been replaced by U. For example, the last two rows in Table (23) for conditionals correspond to the last row F TF in the original Table (18) for conditionals in Section 4.2.4. Thus, the only new thing to remember is where each table puts U.

The appearance of U throughout the tables can be understood as follows. For each binary connective $\star$, there are assignments of truth and falsity to χ, ψ that give $\chi \star \psi$ a truth-value, either T or F. Call such assignments "basic." When either χ or ψ are undefined, we imagine that they might have been filled in with either T or F. Some ways of filling them in would yield a basic assignment that leaves $\chi \star \psi$ with a truth-value. If such filling-in could make $\chi \star \psi$ true but not false then we assign $\chi \star \psi$ truth in this non-basic case as well. If such filling-in could make $\chi \star \psi$ false but not true then we assign $\chi \star \psi$ falsity in this non-basic case. If *both* kinds of basic assignments can be created by filling in U, then $\chi \star \psi$ is left undefined. This idea will become clearer by re-examining each truth table in turn.

First consider Table (20) for negation. Since $\neg$ is unary, we're in a slightly different situation than just described, but the same idea applies. The basic assignments are T to ψ and F to ψ. They make $\neg\psi$ false and true, respectively, as shown in lines 1 and 3 of the table (counting below the bar). We imagine that if ψ is undefined then it might be filled in either with T (which makes $\neg\psi$ false) or with F (which makes $\neg\psi$ true). Since different fill-ins create basic assignments that leave $\neg\psi$ alternatively true and false, we leave $\neg\psi$ undefined if ψ is undefined. Such is the outcome recorded in the second row.

Next consider Table (21) for conjunction. The basic assignments assign T and F to χ and ψ, and yield the familiar results shown in lines 1, 2, 7, and 8 of the table. The rest of the table can be inferred from the basic assignments. If one of χ, ψ is false and the other undefined then we consider the consequences of filling in U with either T or F. We see that the outcome is always a basic assignment in which $\chi \wedge \psi$ is false. Hence, we assign $\chi \wedge \psi$ falsity in this situation. Such reasoning is recorded in rows 5 and 9 of the table. If one of χ, ψ is true and the other undefined then we again consider the consequences of filling in U with either T or F. Now we see that the outcome may be a basic assignment in which $\chi \wedge \psi$ is true (by replacing U with T), or it may be a basic assignment in which $\chi \wedge \psi$ is false (by replacing U with F). Since neither T nor F is uniquely obtained by such filling-in, we leave $\chi \wedge \psi$ undefined in such cases. This reasoning is recorded in rows 3 and 4. Finally, if both χ and ψ are undefined, it is clear that filling in their truth-values can lead to basic assignments in which $\chi \wedge \psi$ is either true or false. So, $\chi \wedge \psi$ is undefined in this case also, as recorded in line 6 of the table.[10]

The same reasoning applies to Table (22) for disjunction. The basic assignments yield lines 1, 2, 7, and 8 of the table. If one of χ, ψ is true and the other undefined then the result of replacing U with either T or F is a basic assignment in which $\chi \vee \psi$ is true. Hence, we assign $\chi \vee \psi$ truth in this situation (lines 3 and 4). If one of χ, ψ is false then the result of filling in for U may be either truth or falsity. Since neither T nor F is uniquely obtained by such filling-in, we leave $\chi \vee \psi$ undefined, as in rows 5 and 9. Finally, if both χ and ψ are undefined, filling in their truth-values can lead alternatively to truth and falsity, yielding line 6.[11]

The truth table (23) for conditionals starts from the hypothesis that $\chi \rightarrow \psi$ is true when both χ and ψ are true, and false when χ is true and ψ false. These are the basic assignments, recorded in rows 1 and 2. Now suppose that χ is undefined and ψ is false. If the U for χ were replaced by T then the conditional would be false; if U were replaced by F then the conditional would still be undefined. Since only F can be realized by filling in U, we assign falsity to this nonbasic case, as shown in row 5.[12] Similar reasoning applies to the case in which χ is undefined and ψ is true; the only basic assignment that can be reached by filling in U assigns truth to the conditional — yielding row 4. If χ is true and ψ undefined then filling in U with T yields a basic assignment of truth to the conditional whereas filling in U with F yields falsity. Since neither truth value is reached uniquely, the conditional is left undefined in this case, as recorded in row 3. Of course, if both ψ and χ have undefined truth-values then, again, both truth values can be reached by filling in the two occurrences of U; hence, row 6 shows the conditional to be undefined in this case. Finally, if

[10]The foregoing developments are familiar from the *supervaluationist* approach to undefined truth-values. For an introduction to the latter theory, see Beall and van Fraassen [10, Ch. 9].

[11]Our interpretation of $\wedge$ and $\vee$ was advanced by de Finetti [23] (although often attributed to Kleene [61]). The tables also appear in McDermott [71] along with alternative tables. McDermott argues that his alternatives capture some plausible readings of conjunction and disjunction in English.

[12]Here we part company with de Finetti and McDermott, who both assume that the truth value of a conditional is undefined if the truth value of either the left hand side or right hand side is undefined.

χ is false then there is no hope of reaching a basic assignment hence no hope of rendering the conditional either true or false; so the conditional is undefined, as recorded in lines 7, 8, 9.

This brings us to Table (24) for biconditionals. The basic assignments are given in rows 1, 2, and 7; that is, we assume that a biconditional is true if its two sides are true, and false if the sides have different truth values. As usual, we infer the rest of the table from the basic assignments. If one side is T and the other U then filling in U can yield basic assignments of either truth or falsity; so this combination is undefined (as in rows 3 and 4). If one side is F and the other U then filling in U can yield a basic assignment of falsity but none of truth; so we get F as in rows 5 and 9. If both sides are false then no basic assignment can be reached, yielding U (row 8). And if both sides are undefined then both kinds of basic assignments can be reached, also yielding U (row 6).

Our assumptions about basic assignments can be summarized as follows.

(27)

$\neg\psi$	true if ψ is false	false if ψ is true
$\chi \wedge \psi$	true if both ψ and χ are true	false if χ is true and ψ false false if χ is false and ψ true false if χ is false and ψ true
$\chi \vee \psi$	true if ψ is true and χ true true if ψ is true and χ false true if ψ is false and χ true	false if χ is false and ψ false
$\chi \rightarrow \psi$	true if χ is true and ψ true	false if χ is true and ψ false
$\chi \leftrightarrow \psi$	true if χ is true and ψ true	false if χ is true and ψ false false if χ is false and ψ true

Using the reasoning rehearsed above, (27) suffices to piece together each of the tables (20) - (24).

10.2.3 Justification

There remains the question of how to justify the tables. Why not choose different tables? Recall that our project is to explain secure inference in the fragment of English that can be naturally translated into $\mathcal{L}$. Call an English argument a "target" if its premises and conclusion have natural translations into $\mathcal{L}$. For target arguments $\{A_1 \cdots A_n\}$ B, we aim to modify Sentential Logic so that $\{A_1 \cdots A_n\} \Rightarrow B$ if and only if $\{\mathsf{trans}(A_1) \cdots \mathsf{trans}(A_n)\} \models \mathsf{trans}(B)$, where $\mathsf{trans}(A_i), \mathsf{trans}(B)$ are the translations of premises and conclusion. In earlier chapters we have considered a wide range of target arguments, some of which struck us as secure others not. Our hope is that these judgments will be predicted by validity within the new logic we are constructing; just the secure arguments will come out valid. Our modifications to Sentential Logic — including the new kind of truth tables seen above — are justified principally by success in getting $\Rightarrow$ aligned with $\models$. That is, the justification for our theory is more *empirical* than *a priori*.

We're not yet in a position to test the coincidence of $\Rightarrow$ and $\models$ because we have not yet defined $\models$ in our new version of Sentential Logic. This will be accomplished shortly. We must also return to the interpretation of secure inference ($\Rightarrow$), however, which was first introduced in Section 1.3. (The symbol $\Rightarrow$ was introduction in Section 8.4.1.) The original interpretation was as follows.

(28) THE "SECURITY" CONCEPT: An argument $A_1 \cdots A_n$ / C is *secure* just in case it is impossible for $A_1 \cdots A_n$ to be true yet C false.

Some reflection about secure inference is advisable because our gappy approach to truth is motivated by the hypothesis that many declarative English sentences (notably, indicative conditionals) are neither true nor false. Indeed, this is our way out of the paradoxical results of Section 8.4 and 8.5, where we seemed to show that indicative conditionals both were and weren't aptly represented by $\rightarrow$ in standard Logic. Both sides of the earlier discussion assumed that indicative conditionals in English are either true or false. Our hope is that giving up this idea will straighten out the correspondence between formal validity and secure inference.

Our original discussion of secure arguments did not envision sentences without truth value. Should such sentences affect our conception of security? Consider again the (overworked) transitivity example:

(29) If Queen Elizabeth dies tomorrow, there will be a state funeral in London within the week. If the sun explodes tomorrow then Queen Elizabeth will die tomorrow. So, if the sun explodes tomorrow, there will be a state funeral in London within the week.

Doesn't it strike you that it is genuinely impossible for both premises of (29) to be true yet the conclusion false? Look, suppose the premises *are* true, and that the sun does explode tomorrow. Then since the second premise is true, Elizabeth dies tomorrow. Because the first premise is true, the funeral will thus take place on schedule as affirmed by the conclusion. Hence, (29) is secure according to the conception (28) of security But you can also see that the argument is no good! Neither of its premises seem false yet the conclusion is not comforted thereby. Whether or not we choose to we call (29) "insecure," the argument seems to be lacking some virtue associated with proper deductive inference.

We are led in this way to replace (28) with a concept that invokes truth gaps; all three of the statements figuring in (29) are without truth-value according to our analysis of conditionals. (After all, the sun is not going to explode tomorrow, and Queen will still be with us.) One way to proceed is via the idea of "partial truth." We say that a set of sentences is *partly true* if none is false and some are true. As a special case, a singleton set $\{C\}$ of sentences is partly true if and only if C is true. Then:

(30) NEW VERSION OF THE "SECURITY" CONCEPT: An argument $A_1 \cdots A_n \,/\, C$ is *secure* (in the new sense) just in case it is impossible for $A_1 \cdots A_n$ to be partly true yet C not to be partly true (that is, impossible for $A_1 \cdots A_n$ to be partly true yet C either false or undefined).

If we analyze English conditionals as suggested by the treatment of $\rightarrow$ in Table (23) (Section 10.2.1) then Definition (30) declares (29) to be insecure, as desired. This is because one possibility is for the queen to die, followed by a funeral but without solar explosion. By Table (23), such a circumstance renders the first premise of (29) true, the second undefined, and the conclusion undefined. Hence, the premises are partly true but the conclusion is not.[13]

[13]The intransitivity of $\rightarrow$ in our new logic will be demonstrated in just this way. See Section

Another reason for accepting (30) as our analysis of "secure argument" in natural language is its analogy to the definition of formal validity within our new logic (to be presented in Section 10.3.2). We hesitate nonetheless to lean too heavily on (30) in what follows. Few people have "partial truth" in mind when they judge the quality of deductive inferences; the felt distinction between deductively good and bad argumentation often seems more immediate and "cognitively impenetrable" than suggested by (30).[14] Some arguments enjoy a perceived, inferential virtue illustrated by clear cases (as in Section 1.3); others seem to lack this virtue. Such judgments are the data against which our logical theory is meant to be tested, much as a theory of syntax is tested against judgments of well-formedness. Success consists in aligning formal validity with the phenomenon of argument security in natural language. If all goes well, arguments that can be revealingly translated into $\mathcal{L}$ will be secure if and only if their translations are formally valid. Success in this enterprise is the principal justification for our truth-tables and other formal maneuvers.[15]

At the same time, our new approach to the truth and falsity of formulas can also be evaluated on more intrinsic grounds. Specifically, we can ask whether the new truth tables seem intuitively sensible as claims about English counterparts to the five connectives of $\mathcal{L}$. The issue is complicated by different attitudes that one can have about the provenance of $\cup$ in the tables. Is it that a given statement fails to have either truth-value, or does it have *both*?[16] Or does the statement have a truth value but we simply don't know it? If the statement is genuinely missing a truth value, is this because of a kind of *category mistake* (as in "Honesty equals its own square root"), or because of vagueness ("Cincinnati is a big city")? Whether a given table fits ordinary usage may depend on how these questions are answered.

To keep things simple, our interpretation of $\cup$ rests on the following policy.

10.4.3 below.

[14]Translation: a mental function is "cognitively impenetrable" if its internal mechanism is inscrutable to introspection; see Fodor (1983) [31].

[15]As usual in this kind of theory-construction, once a successful theory is established, it can be used to adjudicate marginal or unclear cases; the theory can thus become an aid to reason. For discussion, see Goodman [37, §3.2].

[16]This is the opposition between truth "gaps" and truth "gluts." See [10, Ch. 8] for discussion.

We only consider variables with genuine truth values, either true or false. This point is important enough to be framed.

(31) CONVENTION: When choosing English interpretations of the variables of $\mathcal{L}$, we limit our choice to sentences that are either true or false (not both and not neither).

The convention has already governed all our discussion about Sentential Logic. What's noteworthy is that we here reaffirm it for the new logic presently under construction. But now we must ask: where does U come from in our new logic? The answer is that it arises from false left hand sides of conditionals, as seen in the last three rows of Table (23); it also arises from biconditionals with both sides false, as seen in row 8 of Table (24). These are the only situations in which subformulas with defined truth values give rise to formulas with undefined truth value (as can be verified by inspecting the tables).

Now that the origin of U has been nailed down, we can return to the question of whether the new truth tables are intuitively sensible. For negation, conjunction, and disjunction, the new tables seem just as intuitive as the original ones, from standard Sentential Logic. For the new tables agree with the old ones, going beyond them only when one or more constituents has undefined truth value. And in the latter cases, we're confident that the reader will find our choices sensible, even if equally sensible alternatives come to mind.[17]

Table (23) for $\rightarrow$ is of course crucially different from the one offered by standard Sentential Logic. To justify its treatment of conditionals with false antecedents, we cite an experimental study by the psychologist Philip Johnson-Laird [55]. Participants were presented with statements like

> If there is a letter 'A' on the left-hand side of the card then there is a number '3' on the right-hand side.

The task was to examine cards with letters and numbers on various sides. The cards had to be sorted into one of three categories, namely:

[17] For one set of alternatives, see Section 10.6.1 below (and footnote 11, above).

(a) cards truthfully described by the statement (e.g., the one above);

(b) cards falsely described by the statement;

(c) cards to which the statement was irrelevant.

Twenty-four people were tested. Nineteen assigned cards to Category (c) when the card rendered the left hand side of the conditional false. Such responses agree with Table (23).[18] The experiment therefore testifies to the naturalness of denying a truth value to English indicative conditionals with false left hand side.

The treatment of biconditionals in Table (24) is somewhat less intuitive. Specifically, many people find the sentence

> Grass is blue if and only if clouds are green.

to be true, thereby contradicting row 8 of the table. In defense, we observe the following.

(32) FACT: According to Tables (20) – (24), for any assignment of T, F, and U to $\chi, \psi \in \mathcal{L}$, the formulas $(\chi \rightarrow \psi) \wedge (\psi \rightarrow \chi)$ and $\chi \leftrightarrow \psi$ are either both undefined or share the same truth value.

Thus, our tables enforce identity of truth value (or undefinedness) for:

(a) If Smith wins then Jones wins, and if Jones win then Smith wins.

(b) Smith wins if and only if Jones wins.

This appealing outcome supports our treatment of biconditionals.

[18] McDermott [71, p. 1] envisions a test similar to Johnson-Laird's.

10.2.4 New versus standard sentential logic

Let's give a name to the new system of logic that we are presently developing. We'll call the developments based on Definition (19) *Sentential Logic*$_u$ or just *Logic*$_u$. The subscript u reminds us that the new logic involves truth-assignments that are undefined on certain formulas. In contrast, the logic presented in Chapters 3 - 7 will be called *standard Sentential Logic* or just *standard logic*. On some points there is no difference between standard logic and Logic$_u$. The syntax of the two languages is the same; they have the same set $\mathcal{L}$ of formulas. Just the semantics differs. Moreover, the semantics of both logics start from the common idea of a truth-assignment. In each logic, a truth-assignment is a mapping of the variables of $\mathcal{L}$ to $\{\mathsf{T}, \mathsf{F}\}$; variables never have undefined truth values; see Convention (31).

Standard logic and Logic$_u$ diverge only when we extend truth-assignments to complex formulas. The divergence results from the difference between Definition (19), above, and Definition (6) of Section 4.2.2. On a given truth-assignment, Logic$_u$ leaves some formulas with undefined truth value whereas this never happens in standard logic [see Example (25)]. On the other hand, when a formula has defined truth value in both logics, the truth value is the same. That is:

(33) FACT: Let a truth-assignment α and $\varphi \in \mathcal{L}$ be given. Suppose that α gives φ a truth value according to Logic$_u$ [Definition (19)]. Then α gives φ the same truth value according to standard logic [Definition (6) of Section 4.2.2].

Moreover, the semantic difference between Logic$_u$ and standard logic is due only to their respective treatments of conditionals and biconditionals. The following fact puts a sharp point on this observation.

(34) FACT: Suppose that $\rightarrow$ and $\leftrightarrow$ do not occur in $\varphi \in \mathcal{L}$. Then for all truth-assignments α, $\alpha(\varphi)$ in Logic$_u$ is the same as $\alpha(\varphi)$ in standard logic.

Both Facts (33) and (34) are proved using mathematical induction on the number of connectives in a formula.[19]

For another fundamental similarity between the two logics, observe that Logic$_u$ is *truth functional* just like standard logic. That is, two truth-assignments in Logic$_u$ agree about a formula φ if they agree about the variables appearing in φ.[20] See Fact (12) in Section 4.2.3 for more discussion of truth functionality.

Here's a way to understand what's genuinely different between the two logics. Within Logic$_u$, we define the *truth conditions* of $\varphi \in \mathcal{L}$ to be the truth-assignments that make φ true according to (19), and likewise for *falsity conditions*. Just as for standard logic, the truth and falsity conditions of a given formula are disjoint (no truth-assignment makes a formula both true and false). What's different about Logic$_u$ is that truth and falsity conditions for a given formula don't always *exhaust* the set of all truth-assignments.[21] For example, the truth and falsity conditions for $p \to q$ don't include the truth-assignments that make p false (since such truth-assignments are undefined on $p \to q$).

10.2.5 Logic$_u$ versus NTV

Logic$_u$ neither totally embraces nor totally rejects the thesis that conditionals lack truth values. According to Logic$_u$, conditionals have truth values on some truth-assignments but not on others. In this way, we escape the chain of reasoning that started with the claim that if conditionals have truth values then their probabilities must be well defined [see (8) of Section 10.1.2]. From this assumption we were led to contradictory claims about whether the probabilities of conditionals were the corresponding conditional probabilities.

The ability of conditionals in Logic$_u$ to engender U makes their left hand sides function somewhat like *presuppositions*. Roughly, a presupposition of a statement S_1 is a statement S_2 that must be true if S_1 has either truth value T or

[19]For mathematical induction, see Section 2.11.

[20]To agree about variable v, two truth-assignments must assign the same truth value to v.

[21]In other words, truth and falsity conditions in Logic$_u$ are not always a *partition* of the truth-assignments. For explanation of partitions, see Section 2.8. For disjoint sets, see Section 2.6.

F.[22] Such is the case for conditionals within Logic$_u$. A conditional can be either true or false but only if the left hand side is true. There are many constructions in English which possess a truth value contingently upon the truth of another statement, for example:

(35) (a) George W. Bush's doctoral dissertation *A dialectical materialist analysis of Lenin's pledge to leave no child behind* caught everyone by surprise.

(b) The fact that George W. Bush wrote a doctoral dissertation entitled *A dialectical materialist analysis of Lenin's pledge to leave no child behind* was surprising to everyone.

(c) It was surprising that George W. Bush wrote a doctoral dissertation entitled *A dialectical materialist analysis of Lenin's pledge to leave no child behind*.

In each case the presupposition is:

(36) George W. Bush wrote a doctoral dissertation entitled *A dialectical materialist analysis of Lenin's pledge to leave no child behind*.

Unless this sentence is true, none of (35)a-c is either true or false. Logic$_u$ is motivated by the hypothesis that (36) must likewise be true for the conditional

> If George W. Bush wrote a doctoral dissertation entitled *A dialectical materialist analysis of Lenin's pledge to leave no child behind* then everyone was surprised.

to be either true or false.

[22]For a comprehensive introduction to the theory of presupposition, see Beaver [11]. In what follows, we adopt one particular view of presuppositions that is often qualified as "Strawsonian" [98].

10.3 Tautology and validity in Logic$_u$

10.3.1 Tautology in Logic$_u$

To pursue our presentation of Logic$_u$, we need to define the concepts of *tautology* and *validity*. There are two plausible options for the first concept. We could say that:

(37) TENTATIVE DEFINITION: $\varphi \in \mathcal{L}$ is a tautology in Logic$_u$ just in case for all truth-assignments α, $\alpha(\varphi) = \mathsf{T}$.

This definition has the strange consequence, however, that $p \to p$ is not a tautology since $\alpha(p \to p)$ is undefined if $\alpha(p) = \mathsf{F}$. Similarly, $(p \wedge q) \to p$ is not a tautology, etc. The corresponding English sentences (like "If Clinton wins the Marathon then Clinton wins the Marathon") seem guaranteed to be true, so the foregoing definition is a bit off key. We adopt the natural alternative, namely:

(38) TENTATIVE DEFINITION: $\varphi \in \mathcal{L}$ is a tautology in Logic$_u$ just in case for all truth-assignments α, $\alpha(\varphi) \neq \mathsf{F}$.

By $\alpha(\varphi) \neq \mathsf{F}$ we mean that $\alpha(\varphi) = \mathsf{T}$ or $\alpha(\varphi)$ is undefined. According to (38), $p \to p$, $(p \wedge q) \to q$, etc. are Logic$_u$ tautologies, which is an improvement over (37). But things are still not exactly right since (38) implies that there are tautologies whose negations are also tautologies! One such beast is $(p \wedge \neg p) \to q$. For every truth-assignment α, $\alpha(p \wedge \neg p) = \mathsf{F}$, hence $\alpha((p \wedge \neg p) \to q) = \alpha(\neg((p \wedge \neg p) \to q)) = \mathsf{U}$, so neither $(p \wedge \neg p) \to q)$ nor $\neg((p \wedge \neg p) \to q))$ can come out false. This confers tautology status on both. To rid Logic$_u$ of this outrage, we patch up our definition of tautology one last time.

(39) DEFINITION: $\varphi \in \mathcal{L}$ is a tautology in Logic$_u$ just in case:

(a) for all truth-assignments α, $\alpha(\varphi) \neq \mathsf{F}$, and

(b) for some truth-assignment β, $\beta(\varphi) = \mathsf{T}$.

In this case, we write $\models_u \varphi$.

Thus, a Logic$_u$ tautology must be false under no truth-assignments, and true under some. You can see that the added proviso puts both $(p \wedge \neg p) \rightarrow q)$ and $\neg((p \wedge \neg p) \rightarrow q))$ in their place; neither are tautologies. Observe that (39) generalizes the concept of "tautology" in standard logic inasmuch as standard tautologies also are false on no truth-assignments and true on "some" (namely, "all").

Incidentally, notice the little *u* next to $\models$ in the definition. It prevents us from mixing up standard tautologies with tautologies in Logic$_u$. To keep such matters straight, let's record another convention.

(40) CONVENTION: The use of $\models$ presupposes Standard Logic, with truth-values always defined to be either T or F. The use of $\models_u$ presupposes Logic$_u$, with the possibility of undefined truth-values.

Here is an interesting difference between tautology in Logic$_u$ and tautology in standard logic. Suppose that variable p occurs in formula φ, and that $\models \varphi$ (that is, φ is a standard logic tautology). Now replace every occurrence of p in φ by any formula you please, say χ. You must use the same formula χ for all of these replacements, and every occurrence of p (and just these) must be replaced. Call the resulting formula: $\varphi[\chi/p]$.[23] Then $\models \varphi[\chi/p]$, that is, $\varphi[\chi/p]$ is also a tautology of standard logic. For example, replacing every occurrence of p in the tautology $(p \wedge q) \rightarrow p$ by $(r \vee q)$ yields $((r \vee q) \wedge q) \rightarrow (r \vee q)$ which is still a tautology. The latter formula is $(p \wedge q) \rightarrow p[r \vee q/p]$. Let us record the general fact.

(41) FACT: Let $\varphi, \chi \in \mathcal{L}$ and variable p be given. If $\models \varphi$ then also $\models \varphi[\chi/p]$, where $\varphi[\chi/p]$ is the result of replacing each occurrence of p in φ by χ.

We could prove the fact by mathematical induction, but it is perhaps enough to reason as follows. Since φ is a tautology, it is made true by every truth-assignment. In particular, no matter whether a truth-assignment assigns T or F to p, φ comes out true. But every truth-assignment assignment makes χ either true

[23]This notation is intended to have mnemonic value; it means: the formula φ with χ substituted uniformly for p.

or false, so it plays the same role in $\varphi[\chi/p]$ as p plays in φ. Intuitively, it doesn't matter whether p versus χ is the bearer of a truth-value in corresponding spots of φ.

The parallel to Fact (41) does not hold in Logic$_u$. That is:

(42) FACT: There are formulas $\varphi, \chi \in \mathcal{L}$ and variable p such that $\models_u \varphi$ but $\not\models_u\models \varphi[\chi/p]$, where $\varphi[\chi/p]$ is the result of replacing each occurrence of p in φ by χ.

To witness (42), let φ be $p \rightarrow p$. We've seen that $\models_u \varphi$. Let χ be $q \wedge \neg q$. Then $\varphi[\chi/p]$ is $(q \wedge \neg q) \rightarrow (q \wedge \neg q)$. This formula is undefined in every truth-assignment [according to Table (23)], so it fails to meet condition (39)b in our definition of tautology in Logic$_u$.

But we're not sure that this whole business about tautologies matters very much. Our stalking horse is secure inference in English rather than necessary truth. So tautology in Logic$_u$ would be interesting if it were connected to a useful definition of valid argument. The following considerations, however, indicate the contrary.

It is tempting to define validity in terms of tautology like this:

(43) TENTATIVE DEFINITION: Let argument $\varphi_1 \cdots \varphi_n \;/\; \psi$ be given. The argument is valid (in Logic$_u$) just in case $(\varphi_1 \wedge \cdots \wedge \varphi_n) \rightarrow \psi$ is a tautology.

This definition mirrors a corollary of the "Deduction Theorem" discussed in Section 5.2.2 [see Fact (27)]. But it has unwanted consequences. Consider the argument $p \rightarrow q \;/\; (p \wedge r) \rightarrow q$. It comes out valid according to (43) because $(p \rightarrow q) \rightarrow ((p \wedge r) \rightarrow q)$ is a tautology in Logic$_u$. Why is this formula a tautology? Well, the only way for a truth-assignment α to make it false is if $\alpha((p \wedge r) \rightarrow q) = \mathsf{F}$. But in that case $\alpha(p) = \mathsf{T}$ and $\alpha(q) = \mathsf{F}$, so $\alpha(p \rightarrow q) = \mathsf{F}$, which implies [via Table (23)] that $\alpha((p \rightarrow q) \rightarrow ((p \wedge r) \rightarrow q)) = \mathsf{U}$. Hence, no truth-assignment falsifies $(p \rightarrow q) \rightarrow ((p \wedge r) \rightarrow q)$. Moreover, any truth-assignment that renders p, q, r true makes $(p \rightarrow q) \rightarrow ((p \wedge r) \rightarrow q)$ true as well. Thus, the formula meets the conditions stipulated in Definition (39) for being a tautology. Therefore, the tentative Definition (43) declares $p \rightarrow q \;/\; (p \wedge r) \rightarrow q$ to be valid. And this is not what we want! We saw in Section 10.4.3 that arguments of this form do not translate secure inferences of English. We'll review the matter again in Section 10.4.3, below, so won't pause here to resurrect earlier examples. Suffice it to say that this argument (and others that could be cited) reveal the defects in (43). We must frame an alternative definition that avoids endorsing arguments with dubious counterparts in English, while embracing arguments that are genuinely secure.

(44) EXAMPLE: Show that the argument $p \rightarrow q, q \rightarrow r \;/\; p \rightarrow r$ is valid according to (43). We argued in Section 8.5.2 that this argument does not translate a secure inference in English.

10.3.2 Validity in Logic$_u$

To explain our idea about validity, let an argument $\varphi_1 \ldots \varphi_k \;/\; \psi$ be given. Recall from Definition (7) of Section 5.1.2 that in standard logic a truth-assignment α is called "invalidating" just in case α makes each of $\varphi_1 \ldots \varphi_k$ true but ψ false. In Logic$_u$ we loosen this concept as follows.

(45) DEFINITION: Within Logic$_u$, a truth-assignment α is *partially invalidating* for an argument $\varphi_1 \ldots \varphi_k \;/\; \psi$ just in case:

(a) for all $i \in \{1 \ldots k\}$, $\alpha(\varphi_i) \neq \mathsf{F}$;

(b) $\alpha(\psi) \neq \mathsf{T}$;

(c) at least one of $\alpha(\varphi_1) \ldots \alpha(\varphi_k), \alpha(\psi)$ is defined (T or F).

Put differently, α is partially invalidating if it is defined somewhere in $\varphi_1 \ldots \varphi_k$, ψ and can be extended to a function that assigns T to $\varphi_1 \ldots \varphi_k$ and F to ψ. More intuitively, α partially invalidates $\varphi_1 \ldots \varphi_k \;/\; \psi$ in case it looks like an invalidating truth-assignment (in the sense of standard logic) with some (but not all) of its defined values replaced by U.

(46) EXAMPLE: Any truth-assignment α that makes p false and q true is partially invalidating for the argument $p \rightarrow q,\; q \;/\; p$. For, α leaves the

first premise undefined, makes the second true, and the conclusion false. Filling in the U with T makes α look like it assigns truth to both premises, and falsity to the conclusion. The same truth-assignment is partially invalidating for the argument $p \vee q,\ p \rightarrow r\ /\ r$. The truth-assignment that assigns F to all three variables, is not partially invalidating for $\neg q \rightarrow \neg p\ /\ p \rightarrow q$. This is because it assigns U to "all" premises (there's just one) and to the conclusion.

Validity in Logic$_u$ may now be defined in the natural way as the absence of a partially invalidating truth-assignment. That is:

(47) DEFINITION: The argument $\varphi_1 \cdots \varphi_n\ /\ \psi$ is *valid* (in Logic$_u$) just in case there is no partially invalidating truth-assignment for $\varphi_1 \cdots \varphi_n\ /\ \psi$. In this case we write $\varphi_1 \cdots \varphi_n \models_u \psi$. Otherwise, the argument is *invalid* (in Logic$_u$), and we write $\varphi_1 \cdots \varphi_n \not\models_u \psi$.

In standard logic it is similarly the case that an argument is valid just in case there is no invalidating truth-assignment (see Section 5.1.2). What's different for Logic$_u$ is the recourse to *partially* invalidating truth-assignments. Notice again the little *u* next to $\models$ in this definition; it reminds us that the definition has to do with Logic$_u$.

We must ask the same question about Definition (47) as we asked about the new truth tables (20) – (24). Why adopt it? Why not some other definition? As before, the principal justification is that Definition (47) yields close correspondence between $\Rightarrow$ and $\models_u$; it therefore makes validity in Logic$_u$ seem like an explanation of secure inference in English. (Evidence for this claim is presented in Section 10.4.) Also in favor of Definition (47) is that it generalizes the standard account of validity in a natural way (just substituting partially invalidating truth-assignment for the usual, "fully" invalidating truth-assignments).

10.3.3 Remarks about validity in Logic$_u$

Some points about Definition (47) should be brought to light. The first concerns the *Deduction Theorem* for standard Sentential Logic. We repeat it here, from

Section 5.2.2.

(48) FACT: Let $\Gamma \subseteq \mathcal{L}$, and $\varphi, \psi \in \mathcal{L}$ be given. Then in standard Sentential Logic, $\Gamma \cup \{\varphi\} \models \psi$ if and only if $\Gamma \models \varphi \rightarrow \psi$. In particular, if $\Gamma = \emptyset$ then (in standard Sentential Logic), $\varphi \models \psi$ if and only if $\models \varphi \rightarrow \psi$.

For the reasons discussed in Section 10.3.1, we did not define validity from tautology in Logic$_u$. So we have no reason to expect there to be a Deduction Theorem for Logic$_u$. Indeed, we saw earlier that $\models_u (p \rightarrow q) \rightarrow ((p \wedge r) \rightarrow q)$ whereas it is easy to verify that $(p \rightarrow q) \not\models_u ((p \wedge r) \rightarrow q)$ (any truth-assignment that sets p, q to T and r to F is partially invalidating).

Recall that in standard logic, $\models \varphi$ can be understood as $\emptyset \models \varphi$ (this was mentioned in Section 5.2.1). That is, in standard logic, φ is a tautology just in case the "argument" with no premises and φ as conclusion is valid. The same is true in Logic$_u$. From Definition (47), an argument with no premises is valid in Logic$_u$ just in case there is no partial invalidating truth-assignment for $\emptyset$ / φ, and *this* means that $\emptyset \models_u \varphi$ just in case no truth-assignment makes φ false. [A truth-assignment that leaves φ undefined is not partially invalidating since it fails to meet condition (c) of Definition (45).] And the condition that no truth-assignment falsifies φ is just how we defined tautology in Logic$_u$ [see Definition (39)]. To illustrate, no truth-assignment makes the $p \rightarrow p$ false hence $\models_u p \rightarrow p$ and also $\emptyset \models_u p \rightarrow p$.

Next, we note that $\models_u$ is neither strictly stronger nor strictly weaker than $\models$ as a relation between premises and conclusions of arguments. This assertion is illustrated by the following fact.

(49) FACT:

(a) $p \rightarrow \neg q \models_u \neg(p \rightarrow q)$ but $p \rightarrow \neg q \not\models \neg(p \rightarrow q)$

(b) $\neg(p \rightarrow p) \models p$ but $\neg(p \rightarrow p) \not\models_u p$.

To see that $p \rightarrow \neg q \models_u \neg(p \rightarrow q)$, we must consider the cases in which (a) the conclusion is false and (b) it is not defined. If truth-assignment α makes $\neg(p \rightarrow q)$ false then it makes $p \rightarrow q$ true hence it makes p true and q true (here we rely on the fact that α makes p either true or false). But then α makes $p \rightarrow \neg q$ false so α is not invalidating for the inference.[24] If truth-assignment α leaves $\neg(p \rightarrow q)$ undefined then it makes p false [see Table (23), and keep in mind that truth-assignments are always defined on variables]. But then α also leaves $p \rightarrow \neg q$ undefined so again α is not invalidating. This shows that $p \rightarrow \neg q \models_u \neg(p \rightarrow q)$. On the other hand, truth-assignment (e) in Table (18) invalidates $p \rightarrow \neg q \ / \ \neg(p \rightarrow q)$ (in standard logic). So, we've demonstrated (49)a. The first half of (49)b follows from Fact (54) in Section 8.5.5. The second half follows from the invalidating truth-assignment (e) in Table (18), as you can verify. The point of all this is that validity in Logic$_u$ does not guarantee validity in standard logic; and neither does validity in standard logic guarantee validity in Logic$_u$. Neither of $\models_u$ or $\models$ "says more" than the other.

Given the difference between $\models$ and $\models_u$, the rules for making derivations presented in Chapter 6 don't provide insight into validity within Logic$_u$. Recall from Corollary (5) in Section 7.1 that $\models$ and $\vdash$ correspond; an argument is valid if and only if its conclusion can be derived from its premises using the rules of Chapter 6. Since $\models$ and $\models_u$ don't correspond, $\models_u$ and $\vdash$ don't either. So (you ask), what derivational rules correspond to Logic$_u$? Such rules would define a relation $\vdash_u$ of derivation such that:

(50) For all arguments $\varphi_1 \cdots \varphi_n \ / \ \psi$,

$$\{\varphi_1 \cdots \varphi_n\} \models_u \psi \quad \text{if and only if} \quad \{\varphi_1 \cdots \varphi_n\} \vdash_u \psi.$$

Unfortunately, we can't answer your question since no one seems to have presented rules for a derivation relation $\vdash_u$ that satisfies (50). (You can take this fact as a challenge, and an invitation to think about the matter on your own.)

Finally, suppose that $\varphi, \psi \in \mathcal{L}$ are such that $\varphi \models \psi$ and $\psi \models \varphi$. In other words, suppose that φ and ψ are *logically equivalent* in standard Sentential Logic. (We introduced the idea of logical equivalence in Section 5.4.) Then for every truth-assignment α, $\alpha(\varphi) = \alpha(\psi)$. In the same way, suppose that $\varphi \models_u \psi$ and $\psi \models_u \varphi$ in Logic$_u$. It is then also the case that every truth-assignment acts the same way on the two formulas. That is:

[24]For "invalidating truth-assignment," see Definition (7) in Section 5.1.2.

(51) FACT: Let $\varphi, \psi \in \mathcal{L}$ be such that $\varphi \models_u \psi$ and $\psi \models_u \varphi$. Then for every truth-assignment α, either $\alpha(\varphi)$ and $\alpha(\psi)$ are both undefined, or $\alpha(\varphi) = \alpha(\psi)$.

The fact is easily verified from Definition (47). We call such formulas φ, ψ *logically equivalent* in Logic$_u$.

(52) EXERCISE: Demonstrate that for all $\varphi_1 \cdots \varphi_n, \psi \in \mathcal{L}$, $\{\varphi_1 \wedge \cdots \wedge \varphi_n \wedge \psi\} \models_u \psi$.

10.3.4 Assertibility in Logic$_u$

We now consider a property of sentences that is related to their probability.[25] The new property is often called "assertibility." . To begin our reflection, let us ask how probability should be assigned to the following statement.

(53) The first woman to walk on Mars will be American.

The matter is delicate because there may never be a woman who walks on Mars. (Humans might self-destruct long before they get a chance to send women to another planet.) The sensible thing is therefore to consider the chances of (53) *assuming that* some woman walks on Mars. This is tantamount to considering the probability of (53) assuming that (53) has a truth value. So, letting S be (53), we have:

$$\textbf{\textit{Pr}}(S) = \textbf{\textit{Pr}}(S \text{ is true } \mid S \text{ is either true or false}).$$

From our discussion of conditional probability in Section 9.3.6, we see that the foregoing equation implies:

$$\textbf{\textit{Pr}}(S) = \frac{\textbf{\textit{Pr}}(S \text{ is true} \wedge S \text{ is either true or false})}{\textbf{\textit{Pr}}(S \text{ is either true or false})}.$$

[25]The material in this section follows closely the discussion in Mcdermott [71].

Of course:

$$\frac{\boldsymbol{Pr}(S \text{ is true} \wedge S \text{ is either true or false})}{\boldsymbol{Pr}(S \text{ is either true or false})} = \frac{\boldsymbol{Pr}(S \text{ is true })}{\boldsymbol{Pr}(S \text{ is true}) + \boldsymbol{Pr}(S \text{ is false})}$$

So:

$$\boldsymbol{Pr}(S) = \frac{\boldsymbol{Pr}(S \text{ is true })}{\boldsymbol{Pr}(S \text{ is true}) + \boldsymbol{Pr}(S \text{ is false})}$$

Our idea is to extend this analysis of the probability of (53) to formulas of $\mathcal{L}$, where in Logic$_u$ we must similarly deal with absent truth values. Therefore, given $\varphi \in \mathcal{L}$, we define:

(54) $$\boldsymbol{Pr}_u(\varphi) = \frac{\boldsymbol{Pr}(\varphi \text{ is true })}{\boldsymbol{Pr}(\varphi \text{ is true}) + \boldsymbol{Pr}(\varphi \text{ is false})}$$

Here we use ***Pr*** in the sense introduced by Definition (14) in Section 9.3.3; that is, ***Pr*** is a distribution over truth-assignments, and $\boldsymbol{Pr}(\varphi \text{ is true})$ is the sum of the numbers that ***Pr*** assigns to truth-assignments that make φ true. Likewise, $\boldsymbol{Pr}(\varphi \text{ is false})$ is the sum of the numbers that ***Pr*** assigns to truth-assignments that make φ false. $\boldsymbol{Pr}_u$ will be our symbol for *something like* probability in Logic$_u$; it will turn out that $\boldsymbol{Pr}_u$ is not a genuine probability function.

To make (54) a little more explicit, let us write $[\varphi]^t$ to denote the set of truth-assignments that make φ true; we'll use $[\varphi]^f$ to denote the set of truth-assignments that make φ false. Thus, $[\varphi]^t$ and $[\varphi]^f$ are the truth and falsity conditions, respectively, of φ in the sense of Section 10.2.1. These notations apply to Logic$_u$, not to standard logic. Recall from Section 4.4 that $[\varphi]$ is used in standard logic to denote what $[\varphi]^t$ denotes in Logic$_u$. In the earlier context we didn't need to distinguish $[\varphi]^t$ from $[\varphi]^f$ since one was the complement of the other under TrAs; things are not as simple in Logic$_u$. Now we state our official definition of $\boldsymbol{Pr}_u$.

(55) DEFINITION: Let distribution ***Pr*** be given. Then for all $\varphi \in \mathcal{L}$,

$$\boldsymbol{Pr}_u(\varphi) = \frac{\Sigma\{\boldsymbol{Pr}(\alpha) \mid \alpha \in [\varphi]^t\}}{\Sigma\{\boldsymbol{Pr}(\alpha) \mid \alpha \in [\varphi]^t\} + \Sigma\{\boldsymbol{Pr}(\alpha) \mid \alpha \in [\varphi]^f\}}$$

As noted above, *Pr* in this definition is a probability distribution in the sense of Section 9.3.2 [Definition (14)]. That is, *Pr* maps TrAs into a set of nonnegative numbers that sum to one. What differs in the setting of Logic$_u$ is how such distributions are converted into numbers for formulas. [Compare Definition (17) in Section 9.3.3.] Definition (55) can be put more perspicuously in the following notation.

(56) REFORMULATION OF DEFINITION (55): Let distribution *Pr* be given. Then for all $\varphi \in \mathcal{L}$,

$$\boldsymbol{Pr}_u(\varphi) = \frac{\boldsymbol{Pr}([\varphi]^t)}{\boldsymbol{Pr}([\varphi]^t) + \boldsymbol{Pr}([\varphi]^f)}$$

We extend $\boldsymbol{Pr}_u$ to pairs of events by emulating Fact (31) in Section 9.3.6.

(57) DEFINITION: Suppose that *Pr* is a probability distribution over $\mathcal{L}$. Then for all pairs of formulas $\varphi \mid \psi$ such that $\boldsymbol{Pr}_u(\psi) > 0$,

$$\boldsymbol{Pr}_u(\varphi \mid \psi) = \frac{\boldsymbol{Pr}_u(\varphi \wedge \psi)}{\boldsymbol{Pr}_u(\psi)}.$$

If $\boldsymbol{Pr}_u(\psi) = 0$ then $\boldsymbol{Pr}_u(\varphi \mid \psi)$ is not defined.

To keep everything clear, we always use the symbol *Pr* to denote probability in the sense of Section 9.3, in which $\boldsymbol{Pr}(\varphi)$ is obtained by summing the probabilities of the truth-assignments that satisfy φ in the sense of standard logic. We use $\boldsymbol{Pr}_u$ to denote our new-fangled approach to chance in the context of Logic$_u$. The following convention sums up the matter.

(58) CONVENTION: Given a distribution *Pr* that assigns probabilities to truth-assignments, we also use *Pr* to denote the usual extension of the original distribution to formulas and pairs of formulas in $\mathcal{L}$ (as in Section 9.3). We use $\boldsymbol{Pr}_u$ to denote the extension of *Pr* in the sense of Definitions (55) and (57).

Chances calculated using $\boldsymbol{Pr}_u$ have much in common with *Pr*. Notably:

(59) FACT: Let $\boldsymbol{Pr}_u$ be a function of the kind defined in (55). Then there is a distribution $\boldsymbol{Pr}$ of probability such that for all $\varphi \in \mathcal{L}$ that do not contain the symbols $\rightarrow$ and $\leftrightarrow$, $\boldsymbol{Pr}_u(\varphi) = \boldsymbol{Pr}(\varphi)$.

In other words, $\boldsymbol{Pr}_u$ behaves like a genuine probability function except when conditionals (or biconditionals) are involved. When $\rightarrow$ is present, funny things can happen. Suppose that $\varphi \in \mathcal{L}$ is $(p \wedge \neg p) \rightarrow q$. Then $[\varphi]^t = [\varphi]^f = \emptyset$; the truth and falsity conditions are empty. Hence, no matter what the underlying choice of distribution, $\boldsymbol{Pr}_u(\varphi) = \frac{0}{0+0}$. In other words, $\boldsymbol{Pr}_u((p \wedge \neg p) \rightarrow q)$ is not defined. In contrast, $\boldsymbol{Pr}((p \wedge \neg p) \rightarrow q) = 1$ since every truth-assignment satisfies $(p \wedge \neg p) \rightarrow q$ in standard logic. Despite such odd cases, many familiar properties of probability are guaranteed for $\boldsymbol{Pr}_u$, even when $\rightarrow$ is present. For example, the following observation parallels Fact (22) in Section 9.3.4.

(60) FACT: Let $\boldsymbol{Pr}_u$ be a function of the kind defined in (55), and suppose that $\varphi, \psi \in \mathcal{L}$ are logically equivalent in Logic$_u$. Then $\boldsymbol{Pr}_u(\varphi) = \boldsymbol{Pr}_u(\psi)$.

So we see that $\boldsymbol{Pr}_u$ resembles probability, but it is not exactly probability. Since $\boldsymbol{Pr}_u$ is not exactly probability, we must be explicit about its significance. What is it related to? (Genuine probability is related to rational betting ratios; see Section 9.2.4.)

To formulate our claim about $\boldsymbol{Pr}_u$, let a declarative sentence S be given, and consider a person $\mathcal{P}$ engaged in conversation. $\mathcal{P}$ is considering whether to utter S. To decide, $\mathcal{P}$ should evaluate the impact of hearing S on her interlocutor $\mathcal{I}$ (the person with whom $\mathcal{P}$ is conversing). Will hearing S prove useful to $\mathcal{I}$, or will it lead $\mathcal{I}$ to despair? Perhaps $\mathcal{I}$ will be offended by S, or to the contrary find S flattering, or maybe funny, or boring. From this welter of considerations, we wish to isolate just a single concern, whether $\mathcal{P}$ would be *sincere* in uttering S. Put differently, $\mathcal{P}$ would not wish to be guilty of *intentionally misleading* $\mathcal{I}$ by uttering S. We'll say that S is "assertible" to the extent that $\mathcal{P}$ can utter it sincerely, without risk of being intentionally misleading. For example, if the conversation is about basketball, $\mathcal{P}$ might find the following statement to be assertible.

(61) The Celtics are doomed this year (i.e., they won't make the playoffs).

The sentence is assertible because (let us suppose) $\mathcal{P}$ is sincere in her pessimism about the Celtics; so, to utter (61) would not intentionally mislead anyone. Of course, $\mathcal{P}$ might be mistaken! The Celtics might accomplish an astonishing turn-around, and render (61) false. But this would only reveal that $\mathcal{P}$ is poorly informed about basketball, not that she is insincere or intentionally misleading. To be sincere about (61), all that matters is that $\mathcal{P}$ assign the statement sufficiently high probability. So we see that assertibility is connected to personal probability. Now consider the following statement.

(62) If Jason Kidd gets in a slump, the Nets are doomed.

Under what conditions will $\mathcal{P}$ be sincere (not intentionally misleading) in asserting (62)? What seems relevant in this case is the *conditional probability* that the Nets are doomed assuming that Kidd gets into a slump. To the extent that $\mathcal{P}$ thinks this conditional probability is low, she would be insincere and intentionally misleading in asserting (62).

The foregoing discussion is meant to convey the concept of assertibility, but we acknowledge that the matter remains murky.[26] The core idea of sincere utterance (not intentionally misleading) seems nonetheless sufficiently precise to motivate the following claim about the relation of $\boldsymbol{Pr}_u$ to assertibility.

(63) CLAIM: Suppose that English sentence S can be naturally represented by a formula of $\mathcal{L}$. Suppose that distribution $\boldsymbol{Pr}$ over $\mathcal{L}$ corresponds to the beliefs of a person $\mathcal{P}$. Then the assertibility of S for $\mathcal{P}$ is roughly equal to $\boldsymbol{Pr}_u(S)$.

When S does not involve conditionals then (63) reduces to equating assertibility with probability; this is shown by Fact (59). On the other hand, $\boldsymbol{Pr}_u$ and probability diverge in the presence of $\rightarrow$. In this case, (63) equates assertibility with conditional probability, as will be demonstrated shortly.

[26] For more discussion, see [52, Sec. 4.2].

10.4 Consequences of the theory

10.4.1 Claims

So now you've seen our theory of the indicative conditional (built upon a mixture of ideas of previous authors). It attempts to model $\Rightarrow$ with $\models_u$. Parallel to the discussion in Section 8.3, the following criterion of adequacy puts a fine point on this ambition.

(64) CRITERION OF ADEQUACY FOR LOGIC$_u$: For every argument $\varphi_1 \ldots \varphi_n$ / ψ of $\mathcal{L}$, $\varphi_1 \ldots \varphi_n \models_u \psi$ if and only if every argument $P_1 \ldots P_n$ / C of English that is naturally translated into $\varphi_1 \ldots \varphi_n$ / ψ is secure.

Security was discussed in Section 10.2.3. The idea of natural translation into $\mathcal{L}$ was discussed in Section 8.1; it requires that sentences involving conjunctions like "and" be represented by formulas involving $\wedge$, and so forth. Crucially, sentences involving indicative conditionals must be represented using $\rightarrow$ in the more or less obvious way; otherwise the translation does not count as natural. Given such an argument A of English, we claim that A is a secure argument if its natural translation into $\mathcal{L}$ is valid in Logic$_u$. Otherwise, Logic$_u$ is inadequate according to Criterion (64). Conformity with (64) also requires that every invalid argument of Logic$_u$ can be naturally translated into some non-secure argument of English. Finally, going beyond the requirements of (64), we claim that *Pr*$_u$ predicts assertibility, as formulated in (63).

To avoid one source of confusion in the remaining discussion, let us recall how Greek letters are used. When we write $\varphi \vee \psi \not\models \varphi$, for example, we are denying that *every* choice of formula φ, ψ yields a valid argument. It might nonetheless be the case that *some* choice of φ, ψ makes the argument $\varphi \vee \psi / \varphi$ valid. Indeed, letting both φ and ψ be p makes $\varphi \vee \psi / \varphi$ come out valid since it is then $p \vee p / p$.

In this section we consider some predictions of our theory, and try to determine whether they are right or wrong. We concede at the outset that not every nuance of indicative conditionals is predicted by Logic$_u$, even when translations into Logic$_u$ seem to reveal the relevant structure. Consider the following sentences, discussed in Lycan [69, p. 21].

(65) (a) If you open the refrigerator, it will not explode.

(b) If you open the refrigerator then it will not explode.

Lycan claims that the two sentences lend themselves to different uses. The first is reassuring ("Go ahead. Open the refrigerator. It won't explode!") The second provides invaluable information about how to keep the refrigerator from exploding. We agree with these remarkable intuitions, and affirm that nothing in Logic$_u$ accounts for them.[27] Our theory nonetheless gets various other phenomena right. These are examined in the next two subsections (prior to examining some phenomena that are less congenial to the theory).

10.4.2 Nice consequences involving assertibility

Recall from Section 10.1.2 above that the probability of an indicative conditional in English seems to be disconnected from the probability of the corresponding conditional in $\mathcal{L}$. This is because the probability of a sentence of form "If p then q" appears to be the conditional probability of q given p. In contrast, for a wide range of probability distributions ***Pr***, ***Pr***$(p \rightarrow q) \neq$ ***Pr***$(q \mid p)$. For this reason, ***Pr*** seems ill-suited to predicting the assertibility of indicative conditionals. For, we saw in connection with Example (62) in Section 10.3.4 that the assertibility of English conditionals is connected to their conditional probability.

Now let us compute ***Pr***$_u(p \rightarrow q)$. From Table (23) you can see that $[p \rightarrow q]^t = [p \wedge q]$, and $[p \rightarrow q]^f = [p \wedge \neg q]$. It is also clear that for all distributions ***Pr***,

[27]We also have difficulty with the following conditional found as lead sentence in the *New York Times* of April 20, 2004.

> If Madison Avenue is a Frédéric Fekkai lady, groomed and pampered as a best-in-show spaniel, and NoLIta is a fake bohemian with the Yeah Yeah Yeahs on her iPod and a platinum card in her Lulu Guinness bag, then lower Broadway in SoHo is a pastel-clad 13-year-old, giddily in the grip of a sugar rush."

We haven't the foggiest idea what this sentence means.

$\boldsymbol{Pr}([p \wedge q]) + \boldsymbol{Pr}([p \wedge \neg q]) = \boldsymbol{Pr}([p])$ (because $[p \wedge q]) \cap [p \wedge \neg q] = \emptyset$). Combining these facts yields:

$$\boldsymbol{Pr}_u(p \rightarrow q) = \frac{\boldsymbol{Pr}([p \rightarrow q]^t)}{\boldsymbol{Pr}([p \rightarrow q]^t) + \boldsymbol{Pr}([p \rightarrow q]^f)} = \frac{\boldsymbol{Pr}([p \wedge q])}{\boldsymbol{Pr}([p \wedge q]) + \boldsymbol{Pr}([p \wedge \neg q])}$$

$$= \frac{\boldsymbol{Pr}([p \wedge q])}{\boldsymbol{Pr}([p])}.$$

But the latter fraction equals $\boldsymbol{Pr}(p \wedge q)\ /\ \boldsymbol{Pr}(p)$, which is just the conditional probability $\boldsymbol{Pr}(q \,|\, p)$. Summarizing:

(66) FACT: For all probability distributions, ***Pr***,

$$\boldsymbol{Pr}_u(p \rightarrow q) = \boldsymbol{Pr}(q \,|\, p).$$

Putting (66) together with (63), we get the following consequence of Logic$_u$.

(67) CONSEQUENCE: Suppose that distribution *Pr* over $\mathcal{L}$ corresponds to the beliefs of a person $\mathcal{P}$. Then the assertibility of *if-p-then-q* for $\mathcal{P}$ equals $\boldsymbol{Pr}(q \,|\, p)$.

As we observed in Section 10.1.2, (67) seems about right.[28]

We cannot substitute freely for p and q in (66); putting $p \wedge \neg p$ in place of p, for example, yields undefined $\boldsymbol{Pr}_u((p \wedge \neg p) \rightarrow q)$, hence no specific assertibility for sentences that are translated by $(p \wedge \neg p) \rightarrow q)$. Perhaps this is just as well, given the strangeness of, for example:

> If Barbara Bush both does and doesn't vote Republican in 2004 then her son will be reelected president.

For a variety of more reasonable sentences, our theory makes intuitive predictions. Consider:

---

[28]Consequence (67) does not contradict Fact (13) in Section 10.1.2. Fact (13) involves probability *Pr* rather than assertibility Pr_u. The latter function does not meet all the assumptions required of *Pr* in (12).

(68) If Barbara Bush votes Republican then if most voters follow Barbara's lead, her son will be reelected president.

This sentence is naturally represented in $\mathcal{L}$ by $p \rightarrow (q \rightarrow r)$. You can check that it is logically equivalent in Logic$_u$ to $(p \wedge q) \rightarrow r$, which represents:

(69) If Barbara Bush votes Republican and most voters follow Barbara's lead then her son will be reelected president.

The logical equivalence is welcome inasmuch as (68) and (69) seem to express the same idea (as has often been observed). By (60) in Section 10.3.4, both of them are assigned the same assertibility. Moreover, you can easily check that the latter assertibility equals the conditional probability of r given $p \wedge q$. Let's record this fact.

(70) CONSEQUENCE: Suppose that distribution *Pr* over $\mathcal{L}$ corresponds to the beliefs of a person $\mathcal{P}$. Then the assertibility for $\mathcal{P}$ of the English sentences translated by $p \rightarrow (q \rightarrow r)$ and $(p \wedge q) \rightarrow r)$ equals $\boldsymbol{Pr}(r \mid p \wedge q)$.

We take (70) to be another victory for Logic$_u$; the assertibility of the two kinds of sentences does seem to correlate with $\boldsymbol{Pr}(r \mid p \wedge q)$.[29]

Another test of Logic$_u$ concerns the pair $p \rightarrow q$ and $\neg q \rightarrow \neg p$. They are logically equivalent in standard logic, illustrating the principle of *contraposition*. But they are not equivalent in Logic$_u$. Indeed, any truth-assignment that makes both p and q true makes $p \rightarrow q$ true but is undefined on $\neg q \rightarrow \neg p$. According to Logic$_u$, their assertibilities also differ, namely:

$$\boldsymbol{Pr}_u(p \rightarrow q) = \frac{\boldsymbol{Pr}([p \wedge q])}{\boldsymbol{Pr}([p])} = \boldsymbol{Pr}(q \mid p)$$

$$\boldsymbol{Pr}_u(\neg q \rightarrow \neg p) = \frac{\boldsymbol{Pr}([\neg p \wedge \neg q])}{\boldsymbol{Pr}([\neg p])} = \boldsymbol{Pr}(\neg p \mid \neg q)$$

[29]We thus hold Logic$_u$ strictly responsible for embedded conditionals like (68). Other authors (like Adams, 1998) seem more relaxed about the matter.

In Section 10.1.2, above, we noted that $Pr(q \mid p) = Pr(\neg p \mid \neg q)$ is not true in general. Consequently, Logic$_u$ predicts that an English indicative conditional need not have the same assertibility as its contrapositive. To assess the accuracy of this prediction, consider the following conditionals.

(71) (a) If the next prime minister of Britain speaks English then s/he will hail from London. $(p \rightarrow q)$

(b) If the next prime minister of Britain doesn't hail from London then s/he will not speak English. $(\neg q \rightarrow \neg p)$

The unconditional probability that the next British PM will hail from London is reasonably high, hence the conditional probability that this is true given that s/he speaks English is reasonably high. In contrast, the probability that s/he doesn't speak English given that s/he hails from, say, Gloucester is infinitesimal. To our ears, the two conditional probabilities correspond to the respective assertibilities of these sentences. So, the assertibilities are different, as foreseen by Logic$_u$.

The reason the assertibilities are different, of course, is that $p \rightarrow q$ is not equivalent in Logic$_u$ to $\neg q \rightarrow \neg p$. In fact, contrary to standard logic, neither implies the other in Logic$_u$.

10.4.3 Nice consequences involving validity

In Section 8.5 we examined four kinds of arguments that challenge the thesis that *if–then–* is successfully represented by $\rightarrow$ in standard logic. We claim that $\rightarrow$ does better in Logic$_u$.

Transitivity. Transitivity is not in general valid in Logic$_u$. In particular:

(72) FACT: $\{p \rightarrow q, q \rightarrow r\} \not\models_u p \rightarrow r$.

For if a truth-assignment makes p is false and both q and r true then it leaves $p \rightarrow r$ undefined, $p \rightarrow q$ also undefined and $q \rightarrow r$ true. Such a truth-assignment is therefore partially invalidating for the argument $\{p \rightarrow q, q \rightarrow r\} \ / \ p \rightarrow r$

[see Definition (45), above]. Logic$_u$ is thus safe from counter-examples to the transitivity of *if–then–* such as the one discussed in Section 8.5.2. We repeat it here.

(73) If Queen Elizabeth dies tomorrow (q), there will be a state funeral in London within the week (r). If the sun explodes tomorrow (p) then Queen Elizabeth will die tomorrow (q). So, if the sun explodes tomorrow (p), there will be a state funeral in London within the week (r).

Monotonicity. Our new logic certainly does not subscribe to monotonicity:

(74) FACT: $p \rightarrow q \not\models_u (p \wedge r) \rightarrow q$.

After all, if r is false then $(p \wedge r) \rightarrow q$ is undefined. Yet both p, q may be true, rendering $p \rightarrow q$ true as well. Logic$_u$ thus escapes responsibility for the insecure argument presented in Section 10.4.3, namely:

(75) If a torch is set to this very book today at midnight (p) then it will be reduced to ashes by tomorrow morning (q). Therefore, if a torch is set to this very book today at midnight (p) and the book is plunged into the ocean tonight at one second past midnight (r) then it will be reduced to ashes by tomorrow morning (q).

One way or the other. The standard tautology $(p \rightarrow q) \vee (q \rightarrow p)$ is also tautologous in Logic$_u$.

(76) FACT: $\models_u (\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$.

But the definition (39) of tautology in Logic$_u$ renders (76) innocuous, disarming the example given in Section 8.5.4. The example was as follows (for a girl chosen at random from those born in 1850).

(77) At least one of the following statements is true.

If the girl grew up in Naples (p) then she spoke fluent Eskimo (q).

If the girl spoke fluent Eskimo (q) then she grew up in Naples (p).

It isn't the case that (76) commits us to one of $p \rightarrow q$, $q \rightarrow p$ being true. To be a tautology in Logic$_u$ it suffices that not both are false. In particular, if p is true and q false then the first disjunct of $(p \rightarrow q) \vee (q \rightarrow p)$ is false and the second undefined. Hence, neither disjunct is true. The offending (77) is therefore not a consequence of our theory. In terms of validity, the superiority of Logic$_u$ compared to standard logic may be put as follows.

(78) FACT:

(a) $r \models (p \rightarrow q) \vee (q \rightarrow p)$

(b) $r \not\models_u (p \rightarrow q) \vee (q \rightarrow p)$

In other words, in standard logic any sentence (e.g., "Bees sneeze") implies (77) whereas this is not true in Logic$_u$.

Negating conditionals. Logic$_u$ does not validate the passage from $\neg(p \rightarrow q)$ to p (or to $\neg q$). For the record:

(79) FACT: $\neg(p \rightarrow q) \not\models_u p$ [whereas $\neg(p \rightarrow q) \models p$].

For if p is false, $\neg(p \rightarrow q)$ is undefined, yielding a validity-busting transition from undefined to false. Logic$_u$ is therefore immune to the theological example (55) discussed in Section 8.5.5. The latter example was the inference:

> It is not true that if God exists then evil acts are rewarded in Heaven.
> Therefore, God exists.

It is valid when translated into Sentential Logic but not Logic$_u$. Likewise, $\neg(p \rightarrow q) \not\models_u p$ whereas $\neg(p \rightarrow q) \models p$.

We're on a roll! Before the spell is broken, let's examine a few more arguments that are central to the logic of *if–then–*.

Modus Tollens. We mentioned earlier (Section 10.4.2) that Logic$_u$ invalidates contraposition: $p \rightarrow q \not\models_u \neg q \rightarrow \neg p$. For if p, q are both true then $p \rightarrow q$ is true but $\neg q \rightarrow \neg p$ is undefined. On the other hand, Logic$_u$ validates *Modus Tollens*, the inference from $\{p \rightarrow q, \neg q\}$ to $\neg p$.[30] Let us record the contrast.

(80) FACT:

(a) $p \rightarrow q \not\models_u \neg q \rightarrow \neg p$.

(b) $\{p \rightarrow q, \neg q\} \models_u \neg p$.

You can easily verify (80)b. We did not state the latter fact with Greek letters because $\{\varphi \rightarrow \psi, \neg\psi\} \models_u \neg\varphi$ is not true for all formulas φ, ψ [see Exercise (82)]. Fact (80)b brings to mind the earlier example (71) from Section 10.4.2, which concerned (80)a. Here it is again:

(a) If the next prime minister of Britain speaks English then s/he will hail from London. $(p \rightarrow q)$

(b) If the next prime minister of Britain doesn't hail from London then s/he will not speak English. $(\neg q \rightarrow \neg p)$

Can we recast these sentences as a countexample to (80)b? Here goes:

(81) If the next prime minister of Britain speaks English then s/he will hail from London ($p \rightarrow q$). In fact, the next prime minister of Britain will not hail from London ($\neg q$). Therefore, the next prime minister of Britain will not speak English ($\neg p$).

Let us defend the claim that (81) is a secure inference, in accordance with (80)b. First, it seems clear that the conclusion of (81) is either true or false (it does not involve a conditional). Suppose it to be false. Then it is true that the next prime minister of Britain speaks English. Hence, by *Modus Ponens*, if the first premise of (81) is true, it follows that the next prime minister will hail from

[30] *Modus Tollens* was shown earlier to be a valid inference in standard logic. See Fact (6)b of Section 5.1.2.

London.[31] This contradicts the second premise. If the conclusion is false, it is therefore impossible for both premises to be true. Also, supposing that the next prime minister of Britain speaks English makes it implausible that either of the premises could have undefined truth-values. So the argument is secure.

(82) EXERCISE: Produce formulas φ, ψ that show $\{\varphi \rightarrow \psi, \neg\psi\} \models_u \neg\varphi$ not to be true in general.

Modus Ponens. Our defense of (81) relied on *Modus Ponens*. The latter principle warrants the inference from φ and $\varphi \rightarrow \psi$ to ψ. There has been much controversy over the status of this apparently innocuous form of inference (stemming from a provocative paper by McGee [72]). The controversy doesn't affect the use of *Modus Ponens* above, however, since we invoked it in the special case where both φ and ψ are variables; in this case, *if–then–* seems to conform to the principle (at least, no counterexamples have come to anyone's mind). So, we are pleased to record the following fact, which you can easily demonstrate.

(83) FACT: $p \rightarrow q, p \models_u q$

What happens if φ and ψ are logically complex? Is the argument $\varphi \rightarrow \psi, \varphi \;/\; \psi$ valid in Logic$_u$? It turns out to depend on the choice of φ, ψ. We have the following contrast.

(84) FACT:

(a) $(p \rightarrow q) \rightarrow r, p \rightarrow q \models_u r$

(b) $p \rightarrow (q \rightarrow r), p \not\models_u (q \rightarrow r)$

An invalidating truth-assignment for (84)b sets p and r to T, and q to F. The validity claimed in (84)a is verified by marching through the truth-assignments indicated in Table (23).

[31] Reminder: *Modus Ponens* is the inference from *if-φ-then-ψ* and φ to ψ. The formal counterpart $\{\varphi \rightarrow \psi, \varphi\} \;/\; \varphi$ was shown to be valid in standard logic. See Fact (6)a of Section 5.1.2.

We can think of no convincing counterexample to English translation of (84)a so we agree with Logic$_u$ that such arguments are secure (although this may reveal no more than weakness in the authors' imagination!). There remains (84)b. Is there an insecure translation of $p \rightarrow (q \rightarrow r), p \;/\; (q \rightarrow r)$ to English? If not, Logic$_u$ is discredited as a model of *if–then–*.

A non-secure translation of the argument was advanced in McGee [72]. To understand his example, we must review the circumstances of the presidential race of 1980. It featured not one, but two Republicans. There was Ronald Reagan (heavily favored) and John Anderson (not a prayer). Now consider the following argument.

(85) (a) If a Republican won then if Reagan lost then Anderson won.

(b) A Republican won.

(c) *Therefore*: If Reagan lost then Anderson won.

Here, the variables in (84)b have been replaced thusly:

p	:	A Republican won
q	:	Reagan lost
r	:	Anderson won

Many people have the intuition that (85)a,b came true at the close of polling, and (85)c did not. The example has nonetheless divided opinion.[32] For our part, we believe that McGee's example justifies Fact (84)b since the argument he presents is not secure. In support of this conclusion, let us consider a variant of McGee's example, one without the embedded *if–then–* (which might be suspected of clouding our intuitions).

(86) (a) If a Republican won and Reagan lost then Anderson won.

(b) A Republican won.

(c) *Therefore*: If Reagan lost then Anderson won.

[32] Lycan [69, p. 67] agrees that it is a counterexample to *modus ponens* whereas McDermott [71, p. 33], to the contrary, thinks the argument is valid.

Once again, the premises (86)a,b both seem true whereas the conclusion does not. Now, (86)a appears to express just what (85)a expresses, and otherwise the argument (86) is the same as (85). So, the judgment of nonsecurity concerning (86) reinforces our conviction that (85) is also insecure. Of course, this is good news for Logic$_u$ in light of Fact (84)b. Moreover, the apparent equivalence of (86)a and (85)a is another feather in the cap of Logic$_u$ since we have:

(87) FACT: In Logic$_u$, $p \rightarrow (q \rightarrow r)$ is logically equivalent to $(p \wedge q) \rightarrow r$.

More on negating conditionals. What's your view of the following arguments?

(88) (a) It's not the case that if Britney Spears was born in 1900 then she has been straight with the public about her age. *Therefore*, if Britney Spears was born in 1900 then she has not been straight with the public about her age.

(b) If Britney Spears was born in 1900 then she has not been straight with the public about her age. *Therefore*, it's not the case that if Britney Spears was born in 1900 then she has been straight with the public about her age.

If you're like us, you think they are both secure. So, you'll be pleased with the following fact, easily verified via Table (23).

(89) FACT: In Logic$_u$, $\neg(p \rightarrow q)$ and $p \rightarrow \neg q$ are logically equivalent.

Of course, the equivalence does not hold in standard logic. Whereas $\neg(p \rightarrow q) \models p \rightarrow \neg q$, the reverse implication is false.

Other nice facts. Before turning to some bad news, let us record a variety of other reassuring facts about Logic$_u$. Each of the implications that appear in the following table strike us as corresponding to secure arguments in English, or else correspond to arguments that are too strange to engender much intuition either way. Similarly, the nonimplications seem to translate into arguments that are not secure, or at least not clearly secure. We hope you see things the same way.

(90)

$(p \vee q) \rightarrow r, p \models_u r$	$(p \wedge q) \rightarrow r, p \not\models_u r$
$p \rightarrow (q \wedge r), p \models_u q$	$p \rightarrow (q \vee r), p \not\models_u q$
$\neg(p \vee q) \models_u \neg p \wedge \neg q$	$\neg p \wedge \neg q \models_u \neg(p \vee q)$
$\neg(p \wedge q) \models_u \neg p \vee \neg q$	$\neg p \vee \neg q \models_u \neg(p \wedge q)$
$p \wedge (q \vee r) \models_u (p \wedge q) \vee (p \wedge r)$	$(p \wedge q) \vee (p \wedge r) \models_u p \wedge (q \vee r)$
$p \vee (q \wedge r) \models_u (p \vee q) \wedge (p \vee r)$	$(p \vee q) \wedge (p \vee r) \models_u p \vee (q \wedge r)$
$p, q \models_u p \wedge q$	$p \wedge q \models_u p$
$q \not\models_u p \rightarrow q$	$\neg p \not\models_u p \rightarrow q$
$p \wedge \neg p \models_u q$	$\neg(p \rightarrow p) \not\models_u q$
$q \models_u p \vee \neg p$	$q \not\models_u p \rightarrow p$
$p \models_u p \vee q$	$p \rightarrow q \not\models_u q \rightarrow p$
$p \vee q, \neg p \models_u q$	$(p \wedge q) \vee (p \wedge \neg q) \models_u p$
$p \rightarrow (q \wedge \neg q) \models_u \neg p$	$(p \vee \neg p) \rightarrow q \models_u q$

(91) EXERCISE: Demonstrate the claims in Table (90). Which correspond to standard logic? Do they support or infirm Logic$_u$ as a theory of English indicative conditionals?

10.5 Some problematic cases

We'd love to tell you that Logic$_u$ solves the problem of indicative conditionals in natural language. If such were the case, the present authors would already be rich and famous (so wouldn't have bothered to write this book). Alas, Logic$_u$ has some noteworthy defects. Let us face the awful truth.

10.5.1 Inferences from true conditionals

We start with good news. Logic$_u$ avoids validating the following suspicious arguments.

(92) FACT:

(a) $p \rightarrow q \not\models_u p$

(b) $p \rightarrow q \not\models_u q$

To verify (92), suppose in each case that p and q are false. Then each conclusion is false yet the common premise is undefined; such a transition from premise to conclusion yields invalidity in Logic$_u$. It follows that Logic$_u$ avoids declaring secure the inferences from

(93) If dinosaurs invaded Central Park last night then there were a lot of surprised New Yorkers this morning.

to either:

(94) (a) Dinosaurs invaded Central Park last night.

(b) There were a lot of surprised New Yorkers this morning.

Fact (92) is thus reassuring, but it masks a less intuitive feature of our theory. In Logic$_u$, the *truth* of $p \rightarrow q$ guarantees the truth of p and of q. This is an immediate consequence of the truth table (23) of Section 10.2.1, above. So it looks as if Logic$_u$ is committed, after all, to the security of the argument from *if-p-then-q* to p (and to q) — except that the premise of the argument must make it clear that *if-p-then-q* is true, not merely non-false. Yet let us argue in favor of this consequence of Logic$_u$, which we acknowledge is unpalatable at first sight.

If the truth of *if-p-then-q* does not guarantee the truth of p then there must be cases in which *if-p-then-q* is true despite the falsity of p. Imagine, for example, that (93) is true but (94)a is false. To declare (93) true in these circumstances, however, is to say something like: *conditions were right* for a lot of New Yorkers to be surprised this morning *if* (contrary to fact) Dinosaurs invaded Central Park last night. But to affirm that conditions were right in this sense is to take a stand on many issues not evoked by the sentence. For example, it would be required that:

(a) many New Yorkers were in or near Central Park this morning,

(b) the mayor did not issue a dino-alert the day before (in which case New Yorkers would not be surprised),

(c) dinosaur invasions of Central Park are rare events (ditto),

and so forth. A well known theory claims that all (or many) such matters are indeed posed and resolved in determining the truth value of *subjunctive* conditionals.[33] But as discussed in Section 8.2, indicative conditionals are importantly different from the subjunctive kind. We think it plausible that indicative conditionals don't evoke counterfactual possibilities (as subjunctive conditionals clearly do). Thus, when the left hand side is false, the truth of an indicative conditional does not hinge on a myriad of implicit facts, as above. Rather, it simply becomes impossible to evaluate the whole.

In support of this intuition, we have already cited (in Section 10.2.2) the experimental study by Philip Johnson-Laird [55]. We therefore think that *if*-p-*then*-q can't be true unless p is, hence that the truth of *if*-p-*then*-q guarantees the truth of p. It follows that the truth of *if*-p-*then*-q also guarantees that of q, by *Modus Ponens*. As discussed in Section 10.4.3, *Modus Ponens* can be challenged when φ or ψ are logically complex sentences, e.g., themselves involving conditionals. But here we're concerned just with the case in which both φ and ψ are variables, and *Modus Ponens* seems reliable in such circumstances. Let us summarize the fact about Logic$_u$ that prompted the preceding discussion.

(95) FACT: In Logic$_u$, for every truth-assignment α, if $\alpha \in [p \rightarrow q]^t$ then

- (a) $\alpha \in [p]^t$, and
- (b) $\alpha \in [q]^t$.

Fact (95) seems to leave us endorsing inferences like the following.

(96) PREMISE: It is true that if dinosaurs invaded Central Park last night then there were a lot of surprised New Yorkers this morning.

CONCLUSION: It is true that dinosaurs invaded Central Park last night.

[33] See [93, 66].

Doesn't the unacceptability of (96), despite all we have said, count against Logic$_u$ as a model of indicative conditionals? We think not. Logic$_u$ only makes claims about statements that can be naturally translated into $\mathcal{L}$. The connectives of $\mathcal{L}$ represent various uses of "and," "if ... then ...," and so forth. But nothing in $\mathcal{L}$ corresponds to the English expression *It is true that*. Certainly, $\neg$ does not represent this expression; and $\neg$ is the only candidate for the job in $\mathcal{L}$ since it is the only unary connective. (Like $\neg$, "It is true that" attaches to just one sentence at a time.) Hence, the premises and conclusion of (96) cannot be naturally translated into $\mathcal{L}$ in the sense discussed in Section 10.4.1. It follows that Logic$_u$ makes no claim about (96).

Still not convinced? Then let us play our last card. It turns out that Logic$_u$ offers a neat explanation why you might still be tempted to reject the argument (96). By Fact (67), the assertibility of the premise of (96) is the conditional probability that New Yorkers will be surprised *assuming that* dinosaurs invade Central Park. Surely you think that this probability is close to one. On the other hand, the assertibility of the conclusion of (96) corresponds to the probability that dinosaurs invade Central Park, which is no doubt close to zero in your opinion. The difference in the two assertibilities suggests that the perceived non-security of (96) is an illusion based on confusing assertibility for probability of truth. According to Logic$_u$, when conditionals are involved, assertibility is not the same thing as probability of truth. Despite its high assertibility, the probability is quite low that the premise of (96) is true (since the left hand side of the conditional is so likely to be false). The argument seems dubious only because you've let assertibility masquerade as the truth of "if dinosaurs invaded Central Park last night then there were a lot of surprised New Yorkers this morning." Why might you have made this mistake? It's because assertibility and probability-of-true *do* line up for many sentences, namely, the nonconditional ones.

Our first "problem" for Logic$_u$ thus seems like a dud. The next is more worrisome.

10.5.2 Conjunction

One of the most fundamental inferences is from the premises p, q to the conclusion $p \wedge q$. So we may breathe a sigh of relief that $p, q \models_u p \wedge q$. But now observe the following calamity.

(97) FACT: $p, q \rightarrow r \not\models_u p \wedge (q \rightarrow r)$. Thus, it is not generally true that $\varphi, \psi \models_u \varphi \wedge \psi$.

A partially invalidating truth-assignment for the argument $p, q \rightarrow r \ / \ p \wedge (q \rightarrow r)$ assigns truth to p and r, and falsity to q. The conclusion is thus undefined [according to Table (21)] yet one premise is true and the other undefined. By Definition (47), this is enough to show $p, q \rightarrow r \not\models_u p \wedge (q \rightarrow r)$. There is a general lesson to be learned here. A claim like $\varphi, \psi \models_u \varphi \wedge \psi$, written in Greek letters, cannot be inferred from substituting variables for the Greek, as in $p \wedge (q \rightarrow r)$. For Greek letters include the possibility of conditionals, which are undefined on some truth-assignments whereas variables are never undefined.

What's calamitous about (97) is that it predicts the existence of a non-secure English argument that is naturally translated into $p, q \rightarrow r \ / \ p \wedge (q \rightarrow r)$. Absent such an argument, Logic$_u$ falls short of Criterion (64), discussed in Section 10.4.1. And the argument does indeed seem to be absent; at least, our own frantic search has failed to reveal one.

Once a defect as fundamental as (97) shows up, you can be sure that other difficulties are lurking nearby. We were pleased by Fact (72), above, stating that $\rightarrow$ is not transitive in Logic$_u$. But this turns out to be the case only if we don't conjoin the premises in the argument $p \rightarrow q, q \rightarrow r \ / \ p \rightarrow r$. In other words:

(98) FACT: $(p \rightarrow q) \wedge (q \rightarrow r) \models_u p \rightarrow r$ even though $p \rightarrow q, q \rightarrow r \not\models_u p \rightarrow r$.

The validity in Logic$_u$ of $(p \rightarrow q) \wedge (q \rightarrow r) \ / \ p \rightarrow r$ derives from the interaction of our truth-tables for $\wedge$ and $\rightarrow$ in Logic$_u$ [see (21) and (23)]. If a truth-assignment leaves the conclusion of $(p \rightarrow q) \wedge (q \rightarrow r) \ / \ p \rightarrow r$ undefined then it must leave the first conjunct of the premise undefined, hence the entire conjunction undefined; and if a truth-assignment makes the conclusion false then it must make r false, hence make one of the two conjuncts false (and thus the entire conjunction false).

The same problem besets other arguments that we want Logic$_u$ to declare invalid. Thus, Example (85) in Section 10.4.3 motivates the invalidity of $p \rightarrow (q \rightarrow r), p \ / \ (q \rightarrow r)$ [see Fact (84)b]. Yet conjoining the premises yields the valid $(p \rightarrow (q \rightarrow r)) \wedge p \ / \ (q \rightarrow r)$ (as you can verify).

Before contemplating potential solutions to our woes, let us note that they extend to disjunction because of the logical equivalence in Logic$_u$ of $(p \wedge q)$ and $\neg(\neg p \vee \neg q)$ (easily checked). Thus, from Fact (97) we also derive the unpalatable:

(99) FACT: $p, q \rightarrow r \not\models_u \neg(\neg p \vee \neg(q \rightarrow r))$. Thus, it is not generally true that $\varphi, \psi \models_u \neg(\neg\varphi \vee \neg\psi)$.

(100) EXERCISE: Does $\varphi \wedge \psi \models_u \varphi$ hold?

10.6 Can our theory be repaired?

10.6.1 A new logic

The difficulties described in Section 10.5.2 suggest revision of Tables (21) and (22) for conjunction and disjunction in Logic$_u$. Suppose that truth-assignment α makes χ true but leaves ψ undefined. Then according to Logic$_u$, α also leaves the conjunction $\chi \wedge \psi$ undefined. This is because *if ever α were to be defined on ψ*, the truth value of $\chi \wedge \psi$ would depend on whether $\alpha(\psi)$ were true or false. In contrast, if α makes χ false (but still leaves ψ undefined) then α makes $\chi \wedge \psi$ false. This is because *even if α were ever defined on ψ*, α would assign falsity to $\chi \wedge \psi$. Such is the idea behind Definition (19) (See Section 10.2.1).

Now that things are turning out badly for Logic$_u$, the time has come to heap abuse on this idea. (We didn't dare do so while Logic$_u$ looked like a winner.) What on earth does it mean to contemplate the possibility that α might *one day* be defined on ψ even though *today* it is not? Whether α is defined on ψ is an (eternal) mathematical fact, not the kind of thing that changes. Compare: "$2/(3^2 - 9)$ is undefined but *if ever* 3^2 were to equal 10 then $2/(3^2 - 9)$ would equal 2." Rather than defending such discourse, let us try a different rationale for assigning truth values to conjunctions.

Suppose as before that α makes χ true but leaves ψ undefined. Then we might say:

> "It makes no sense waiting around for α to be defined on ψ; it just isn't and never will be. So let's work with what we have in hand. The conjunction $\chi \wedge \psi$ offers just one conjunct with defined truth-value, and the value is T. Since this is the only indication of truth-value that we have, we'll generalize it to the whole conjunction, declaring α to make the conjunction T as well."

In the same way, if α makes χ false (still leaving ψ undefined) then α should decide about $\chi \wedge \psi$ using just χ, which points to F for the conjunction. Finally, if α is undefined on *both* χ and ψ then α has no information to guide it, so must leave $\chi \wedge \psi$ undefined. The upshot of this reasoning is a new truth table for conjunction, as follows. It differs from Table (21) just in rows 3 and 4.

(101) YET ANOTHER TABLE FOR CONJUNCTION:

χ	$\wedge$	ψ
T	T	T
T	F	F
T	T	U
U	T	T
U	F	F
U	U	U
F	F	T
F	F	F
F	F	U

The same reasoning applies to disjunction. Suppose that α is defined on χ but not on ψ. Then α should treat $\chi \vee \psi$ according to the sole disjunct in play, assigning T to the disjunction if T was assigned to χ and F otherwise. We thus obtain the following, new truth table for disjunction. It differs from Table (22) in rows 5 and 8.[34]

(102) YET ANOTHER TABLE FOR DISJUNCTION:

χ	$\vee$	ψ
T	T	T
T	T	F
T	T	U
U	T	T
U	F	F
U	U	U
F	T	T
F	F	F
F	F	U

The foregoing tables are the sole modifications we propose for our theory. The balance of Definition (19) is unchanged, as is Definition (45) of "partially invalidating" truth-assignment, and Definition (47) of validity. The replaced clauses of Definition (19) conform to Tables (101) and (102), hence read as follows.

(103) DEFINITION: Suppose that a truth-assignment α and a formula φ are given, where φ is either a conjunction or a disjunction.

(c) Suppose that φ is the conjunction $\chi \wedge \psi$. Then $\alpha(\varphi) = \mathsf{T}$ just in case (a) $\alpha(\chi) = \mathsf{T}$ and $\alpha(\psi) = \mathsf{T}$, (b) $\alpha(\chi) = \mathsf{T}$ and $\alpha(\psi)$ is undefined, or (c) $\alpha(\chi)$ is undefined and $\alpha(\psi) = \mathsf{T}$. If either $\alpha(\chi) = \mathsf{F}$ or $\alpha(\psi) = \mathsf{F}$, then $\alpha(\varphi) = \mathsf{F}$. In the one other case, $\alpha(\varphi)$ is not defined.

(d) Suppose that φ is the disjunction $\chi \vee \psi$. Then $\alpha(\varphi) = \mathsf{F}$ just in case (a) $\alpha(\chi) = \mathsf{F}$ and $\alpha(\psi) = \mathsf{F}$, (b) $\alpha(\chi) = \mathsf{F}$ and $\alpha(\psi)$ is undefined, or (c) $\alpha(\chi)$ is undefined and $\alpha(\psi) = \mathsf{F}$. If either $\alpha(\chi) = \mathsf{T}$ or $\alpha(\psi = \mathsf{T}$, then $\alpha(\varphi) = \mathsf{T}$. In the one other case, $\alpha(\varphi)$ is not defined.

For the revised system, let us use the exciting new name Logic^u. (The placement of the u suggests that Logic^u has the upper hand on Logic$_u$.) Of course, we'll need to supplement Convention (40) with this one:

[34]The new tables for conjunction and disjunction are foreseen in McDermott [71, p. 5] as variants of the original tables, representing aspects of the usage of "and" and "or" in English.

(104) CONVENTION: The use of $\models$ presupposes Standard Logic, with truth-values always defined to be either T or F. The use of $\models^u$ presupposes Logic^u, with the possibility of undefined truth-values.

So now we have a new logic, based on a new idea about conjunction and disjunction. Does it resolve the problems that plague Logic$_u$?

(105) EXERCISE: Show that:

(a) Both $p \rightarrow (p \vee q)$ and $\neg p \rightarrow (\neg p \vee q)$ are tautologies in Logic$_u$.

(b) Both $p \rightarrow (p \vee q)$ and $\neg p \rightarrow (\neg p \vee q)$ are tautologies in Logic^u.

(c) The conjunction of $p \rightarrow (p \vee q)$ and $\neg p \rightarrow (\neg p \vee q)$ is not a tautology in Logic$_u$.

(d) The conjunction of $p \rightarrow (p \vee q)$ and $\neg p \rightarrow (\neg p \vee q)$ *is* a tautology in Logic^u.

More generally, show that the set of tautologies in Logic^u is *closed under conjunction*, that is, if $\varphi, \psi \in \mathcal{L}$ are Logic^u tautologies then so is $\varphi \wedge \psi$. Do these facts point to an advantage of Logic^u over Logic$_u$? (This exercise is derived from a discussion in Edgington [28].)

10.6.2 Evaluating the modified system

The answer is that Logic^u gives the right answer where Logic$_u$ erred but Logic^u makes a few new mistakes. Let's consider the good news first.

Recall from Fact (97) that $p, q \rightarrow r \not\models_u p \wedge (q \rightarrow r)$, so it it not generally true that $\varphi, \psi \models_u \varphi \wedge \psi$. In contrast:

(106) FACT: $p, q \rightarrow r \models^u p \wedge (q \rightarrow r)$. More generally, $\varphi, \psi \models^u \varphi \wedge \psi$.

As a consequence, Logic^u offers the same judgment about the transitivity of $\rightarrow$ whether or not the premises are conjoined. That is:

(107) FACT: $(p \rightarrow q) \wedge (q \rightarrow r) \not\models^u p \rightarrow r$, just as $p \rightarrow q, q \rightarrow r \not\models^u p \rightarrow r$.

Similarly, we get the same (desired) answer about Example (85) in Section 10.4.3. In Logic^u, both $p \rightarrow (q \rightarrow r), p \ / \ (q \rightarrow r)$ and $(p \rightarrow (q \rightarrow r)) \wedge p \ / \ (q \rightarrow r)$ are invalid (as you can verify).

The other nice consequences of Logic$_u$ explained in Section 10.4.3 are also preserved in Logic^u. Parallel to the facts seen earlier, we have:

(108) FACT:

(a) $p \rightarrow q \not\models^u (p \wedge r) \rightarrow q$.

(b) $\neg(p \rightarrow q) \not\models^u p$

(c) $p \rightarrow q \not\models^u \neg q \rightarrow \neg p$.

(d) $\{p \rightarrow q, \neg q\} \models^u \neg p$.

(e) $\{p \rightarrow q, p\} \models^u q$

(f) $(p \rightarrow q) \rightarrow r, p \rightarrow q \models^u r$

(g) In Logic^u, $\neg(p \rightarrow q)$ and $p \rightarrow \neg q$ are logically equivalent.

As a bonus, $(p \rightarrow q) \vee (q \rightarrow p)$ is not a tautology in Logic^u; it is falsified by assigning T to p and F to q. [Compare Fact (76).]

Since disjunction was modified in Table (102) just as conjunction was modified in Table (101), the two connectives are harmoniously related in Logic^u, just as before. Thus, we have the following contrast to Fact (99).

(109) FACT: $p, q \rightarrow r \models^u \neg(\neg p \vee \neg(q \rightarrow r))$. It is generally the case that $\varphi, \psi \models^u \neg(\neg\varphi \vee \neg\psi)$.

Finally, we claim (without presenting the tedious proofs) that all the nice facts cited in Table (90) for Logic$_u$ remain true for Logic^u. That is:

(110)

$(p \vee q) \rightarrow r, p \models^u r$	$(p \wedge q) \rightarrow r, p \not\models^u r$
$p \rightarrow (q \wedge r), p \models^u q$	$p \rightarrow (q \vee r), p \not\models^u q$
$\neg(p \vee q) \models^u \neg p \wedge \neg q$	$\neg p \wedge \neg q \models^u \neg(p \vee q)$
$\neg(p \wedge q) \models^u \neg p \vee \neg q$	$\neg p \vee \neg q \models^u \neg(p \wedge q)$
$p \wedge (q \vee r) \models^u (p \wedge q) \vee (p \wedge r)$	$(p \wedge q) \vee (p \wedge r) \models^u p \wedge (q \vee r)$
$p \vee (q \wedge r) \models^u (p \vee q) \wedge (p \vee r)$	$(p \vee q) \wedge (p \vee r) \models^u p \vee (q \wedge r)$
$p, q \models^u p \wedge q$	$p \wedge q \models^u p$
$q \not\models^u p \rightarrow q$	$\neg p \not\models^u p \rightarrow q$
$p \wedge \neg p \models^u q$	$\neg(p \rightarrow p) \not\models^u q$
$q \models^u p \vee \neg p$	$q \not\models^u p \rightarrow p$
$p \models^u p \vee q$	$p \rightarrow q \not\models^u q \rightarrow p$
$p \vee q, \neg p \models^u q$	$(p \wedge q) \vee (p \wedge \neg q) \models^u p$
$p \rightarrow (q \wedge \neg q) \models^u \neg p$	$(p \vee \neg p) \rightarrow q \models^u q$

So that's the good news about Logic^u. Now here's the bad news.

(111) FACT:

(a) $p \wedge (q \rightarrow r) \not\models^u q \rightarrow r$. Thus, it is not generally true that $\chi \wedge \psi \models^u \psi$.

(b) $p \rightarrow q \not\models^u r \vee (p \rightarrow q)$. Thus, it is not generally true that $\chi \models^u \psi \vee \chi$.

To verify (111)a, suppose that truth-assignment α makes p, r true and q false. Then α is undefined on $q \rightarrow r$, and makes $p \wedge (q \rightarrow r)$ true; α is therefore partially invalidating for $p \wedge (q \rightarrow r)$ / $q \rightarrow r$. To verify (111)b, suppose that truth-assignment β makes p, q, r false. Then β is undefined on $p \rightarrow q$, and makes $r \vee (p \rightarrow q)$ false; true; β is therefore partially invalidating for $p \rightarrow q$ / $r \vee (p \rightarrow q)$. These results are unwelcome if you accept the security of arguments like the following.

(112) (a) Robins are birds, and if grass is red then it is also blue. *Therefore,* if grass is red then it is also blue.

(b) If grass is red then it is also blue. *Therefore,* either lions are fish or if grass is red then it is also blue.

Sometimes we can almost reconcile ourselves to rejecting the security of (112)a,b. This attitude rests on the observation that reasoning to or from conditionals with false left hand side is apt to be confusing. Mostly, however, we regret Fact (111), and suspect that it signals a telling defect in Logic^u. This difficulty appears to stem from the idea of a partially invalidating truth-assignment, and its role in defining validity.[35] Only these ideas are responsible for the following, depressing observation.

(113) FACT: $p, q \rightarrow r \not\models^u q \rightarrow r$. Hence, $\varphi, \psi \models^u \psi$ is not generally true. Likewise, in Logic$_u$, $\varphi, \psi \models_u \psi$ is not generally true

This last disaster might be addressed with a (yet) more complicated definition of validity. It would go something like this.

(114) DEFINITION: The argument $\varphi_1 \cdots \varphi_n \mathbin{/} \psi$ is *valid in the subset sense* just in case there is some subset S of $\{\varphi_1 \cdots \varphi_n\}$ such that $S \mathbin{/} \psi$ is valid in the original sense of Definition (47) — that is, such that there is no partially invalidating truth-assignment for $S \mathbin{/} \psi$.

Then we get:

(115) FACT: In both Logic$_u$ and Logic^u, the argument $\varphi, \psi \mathbin{/} \psi$ is valid in the subset sense.

But Definition (114) does not address the problem embodied in Fact (111). And no easy fix comes to mind.

(116) EXERCISE: As in Exercise (105), let $\varphi, \psi \in \mathcal{L}$ be $p \rightarrow (p \vee q)$ and $\neg p \rightarrow (\neg p \vee q)$, respectively. Show that $\varphi \wedge \psi \not\models^u \varphi$. This is another illustration of the difficulty signaled in (111)a.

[35] See Definitions (45) and (47), and remember that they have been transported intact from Logic$_u$ to Logic^u.

10.7 Fare thee well

What now? It seems that the options are as follows.

(a) We can fiddle some more with the truth-tables and the definition of partially invalidating truth-assignment, hoping to find some combination that gets everything right. This option retains our judgments about which English arguments are secure or not, and tries to adjust the relation of validity accordingly.

(b) We can search for reasons to change our minds about what we take to be secure and non-secure arguments in English. This option tries to alter what counts as the "correct" concept of validity in our logic.

(c) We can become more modest in our aspirations and exclude from consideration most contexts involving embedded conditionals. Specifically, we might only allow embedding within the two contexts $\neg(\varphi \rightarrow \psi)$ and $\varphi \rightarrow (\psi \rightarrow \theta)$. The first would be treated as equivalent to the non-embedded conditional $\varphi \rightarrow \neg\psi$ and the latter as equivalent to $(\varphi \wedge \psi) \rightarrow \theta$. This option is motivated by the fact that all the problems for Logic_u and Logic^u involve embedded conditionals.[36]

(d) We can pursue an entirely different approach to understanding indicative conditionals in English, perhaps one that is not truth functional. As discussed in Section 10.2.4, Logic_u is truth functional, and the same can be said of Logic^u. One family of approaches along this line considers different kinds of semantic evaluation (as in Lycan [69] or Stalnaker [93, Ch. 7]). Another focuses exclusively on assertibility and does not assign semantic values.

(e) Another approach is to explore the idea (introduced in Section 10.1.1) that *if–then–* does not function as a connective in English, at least, not like the connectives "and" and "or." On this view, an *if–then–* sentence does not make an unqualified assertion, but rather makes a conditional assertion of the right hand side provided that the left hand side turns out to be true.

[36] Approaches to conditionals that limit embedding are developed in [4] and [73].

> If the left hand side turns out to be false, nothing has been asserted. (This idea was advanced by Quine [82, p. 21].)

You can think of (a) as finding a better solution to the original problem, preserving the original constraints and the data. Option (b) changes the data but keeps the problem and constraints. Option (c) changes the problem by limiting it, while Option (d) modifies the constraints on a solution. Option (e) also seems to modify the constraints on a solution to the problem.

For our part, we have pursued (a) rather systematically, without doing better than Logic$_u$ and Logic^u.[37] We remain open-minded about (b) but have yet to be moved from the judgments presented in this book. Option (c) seems to surrender too much territory to the enemy. Option (e) makes it hard to understand embedded conditionals, like:

> If if the bulb is planted, it will become a tulip, then if the acorn is planted, it will become a mighty oak.

There remains (d), which leads to the thought that indicative conditionals in English might not be cleanly separable from subjunctive conditionals — despite our attempt to so separate them in Section 8.2.[38]

And now, dear friends, having presented some options for further investigation, we must take our leave. It's been a great pleasure exploring with you the issues surrounding the relation between formal logic and natural language. We hope that the interest and complexity of these matters has been rendered vivid by the preceding ten chapters. If just this much has been accomplished then the present authors will feel their labor to be amply rewarded.

---

[37] In particular, we're not favorable to the de Finetti/McDermott tables for $\rightarrow$ because they validate transitivity along with the arguments $r\ /\ p \rightarrow q) \rightarrow p$ and $r\ /\ p \rightarrow q) \rightarrow q$. English counterparts of the latter strike us as strange (examples left for you!). You've already seen our objection to transitivity.

[38] Nontruth functional accounts of conditionals include [25], and those discussed in [45, 78], and [69, 10].

Bibliography

Bibliography

[1] *The American Heritage Dictionary of the English Language (Fourth Edition)*, 2000.

[2] D. Adams. *The Hitchhikers Guide to the Galaxy*. Ballantine, New York NY, 1979.

[3] E. Adams. *The Logic of Conditionals*. Reidel, Dordrecht, 1975.

[4] E. Adams. *A Primer of Probability Logic*. Number 68. CSLI Lecture Notes, Stanford CA, 1998.

[5] E. W. Adams. On the logic of conditionals. *Inquiry*, 8:166–197, 1965.

[6] E. W. Adams. Subjunctive and indicative conditionals. *Foundations of Language*, 6:89 – 94, 1970.

[7] A. R. Anderson. A Note on Subjunctive and Counterfactual Conditionals. *Analysis*, 12:35–38, 1951.

[8] Antoine Arnauld and Pierre Nicole. *Logic or the Art of Thinking (edited by Jill Vance Buroker)*. Cambridge University Press (1996), Cambridge, England, 1662.

[9] J. Baron. *Thinking and Deciding (Third Edition)*. Cambridge University Press, New York, 2000.

[10] J. C. Beall and B. C. van Fraassen. *Possibilities and Paradox: an introduction to modal and many-valued logic*. Oxford University Press, New York NY, 2003.

[11] D. I. Beaver. Presupposition. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*, pages 939 – 1008. Elsevier, Amsterdam, 1997.

[12] J. Bennett. *A Philosophical Guide to Conditionals*. Oxford, Oxford UK, 2003.

[13] R. Bradley. More triviality. *Journal of Philosophical Logic*, 28(2):129–139, 1999.

[14] R. Carston. Conjunction and pragmatic effects. *Encyclopedia of Language and Linguistics*, 1993.

[15] E. Castillo, J. M. Gutiérrez, and A. S. Hadi. *Expert Systems and Probabilistic Network Models*. Springer, New York NY, 1997.

[16] Gennaro Chierchia and Sally McConnell-Ginet. *Meaning and Grammar: An Introduction to Semantics (2nd Edition)*. MIT Press, Cambridge MA, 2000.

[17] Charles S. Chihara. The Howson-Urbach proofs of Bayesian principles. In E. Eels and B. Skyrms, editors, *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge University Press, New York NY, 1994.

[18] N. Chomsky. *Syntactic Structures*. Mouton, Amsterdam, 1957.

[19] N. Chomsky. *Language and Mind (2nd Edition)*. Harcourt Brace Jovanovich, New York NY, 1972.

[20] L. J. Cohen. *An Introduction to the Philosophy of Induction and Probability*. Oxford University Press, Oxford, UK, 1989.

[21] L. Cooper. *Aristotle, Galileo, and the tower of Pisa*. Cornell University Press; London, H. Milford, Oxford University Press, Ithaca, NY, 1935.

[22] B. Dancygier. *Conditionals and Prediction*. Cambridge University Press, New York NY, 1998.

[23] B. de Finetti. La logique de la probabilité. In *Actes du Congrès Internationale de Philosophie Scientifique*, volume 391 of *Actualités Scientifiques et Industrielles*, pages 31 – 39, Paris, 1935. Sorbonne IV Induction et probabilité, Hermann et Cle, Editeurs. Translated by Brad Angell as 'The logic of probability,' in *Philosophical Studies* 77, 1995, pp. 181-190.

[24] A. DeMorgan. *Formal Logic*. Taylor, London, 1847.

[25] V. H. Dudman. On conditionals. *The Journal of Philosophy*, 91(3):113 – 128, 1994.

[26] H. D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical Logic, Second Edition*. Springer-Verlag, Berlin, 1994.

[27] D. Edgington. Do Conditionals Have Truth Conditions? In F. Jackson, editor, *Conditionals*. Oxford University Press, Oxford England, 1991.

[28] D. Edgington. Conditionals. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Fall 2001.

[29] H. Enderton. *A Mathematical Introduction to Logic (2nd Edition)*. Academic Press, San Diego, 2000.

[30] Yael Abarbanel-Vinov et al. On the Effective Deployment of Functional Formal Verification. *Formal Methods in System Design*, 19:35–44, 2001.

[31] J. A. Fodor. *The modularity of mind*. MIT Press, Cambridge MA, 1983.

[32] Robert Freidin. *Foundations of Generative Syntax*. MIT Press, Cambridge MA, 1992.

[33] Dov M. Gabbay and F. Guenthner, editors. *Handbook of Philosophical Logic, 2nd Edition*, volume 7. Kluwer Academic Publishers, Dordrecht, 2002.

[34] Thomas Gilovich, Dale Griffin, and Daniel Kahneman, editors. *Heuristics and Biases: The Psychology of Intuitive Judgment*. Cambridge University Press, New York City, 2002.

[35] C. Glymour. *Thinking Things Through*. MIT Press, Cambridge MA, 1992.

[36] N. Goodman. *The Structure of Appearance (Third Edition)*. D. Reidel, Boston MA, 1977.

[37] N. Goodman. *Fact, Fiction and Forecast (4th Edition)*. Harvard University Press, Cambridge MA, 1983.

[38] W. Gustason. *Reasoning from Evidence: Inductive Logic*. Macmillan, New York NY, 1994.

[39] I. Hacking. *An Itroduction to Probability and Inductive Logic*. Cambridge University Press, Cambridge UK, 2001.

[40] T. Hailperin. *Boole's Logic and Probability (2nd edition)*. North Holland, Amsterdam, 1986.

[41] T. Hailperin. *Sentential Probability Logic*. Lehigh University Press, Bethlehem PA, 1996.

[42] J. Halpern, R. Harper, N. Immerman, P. Kolaitis, M. Vardi, and V. Vianu. On the unusual effectiveness of logic in computer science. *Bulletin of Symbolic Logic*, 7(2):213–236, 2001.

[43] J. Y. Halpern. *Reasoning about Uncertainty*. MIT Press, Cambridge MA, 2003.

[44] W. L. Harper. A sketch of some recent developments in the theory of conditionals. In W. L. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, pages 3–38. D. Reidel, Dordrecht, Holland, 1980.

[45] W. L. Harper, R. Stalnaker, and G. Pearce, editors. *Ifs*. D. Reidel, Boston MA, 1981.

[46] Reid Hastie and Robyn M. Dawes. *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*. Sage, London, England, 2001.

[47] J. Hintikka. *The Principles of Mathematics Revisited*. Cambridge University Press, New York NY, 1998.

[48] Wilfred Hodges. *Logic*. Penguin Books, New York NY, 1977.

[49] Whitney Houston. *Love Will Save The Day*. Arista Records Inc., New York, NY, 1987.

[50] C. Howson and P. Urbach. *Scientific Reasoning: The Bayesian Approach (2nd Edition)*. Open Court, La Salle, Illinois, 1993.

[51] F. Jackson. On assertion and indicative conditionals. *Philosophical Review*, 88:565–589, 1979.

[52] F. Jackson. *Mind, Method and Conditionals*. Routledge, London UK, 1998.

[53] Richard C. Jeffrey. *The Logic of Decision (2nd Edition)*. The University of Chicago Press, Chicago IL, 1983.

[54] Richard C. Jeffrey. *Formal Logic: Its Scope and Limits*. McGraw-Hill, New York NY, 1991.

[55] P. N. Johnson-Laird and J. Tagart. How implication is understood. *American Journal of Psychology*, 82:367–73, 1969.

[56] C. G. Jung. *Man and his symbols*. Doubleday, Garden City, 1964.

[57] Daniel Kahneman and Amos Tversky. Choices, Values and Frames. In *Choices, Values and Frames*. Cambridge University Press, New York NY, 2000.

[58] Jerrold J. Katz. *Semantic Theory*. Harper & Row, New York, 1972.

[59] P. Kay and W. Kempton. What is the sapir-whorf hypothesis? *American Anthropologist*, 86:65–79, 1984.

[60] Richard L. Kirkham. *Theories of Truth: A Critical Introduction*. MIT Press, Cambridge MA, 1992.

[61] S. C. Kleene. *Introduction to Metamathematics*. North Holland, Amsterdam, 1952.

[62] S. C. Kleene. *Mathematical Logic*. John Wiley & Sons, New York NY, 1967.

[63] Richard Larson and Gabriel Segal. *Knowledge of Meaning*. MIT Press, 1995.

[64] H. Lasnik. The Forms of Sentences. In L. Gleitman and M. Liberman, editors, *Invitation to Cognitive Science: Thinking (Second Edition)*. M.I.T. Press, Cambridge MA, 1995.

[65] S. Levinson. Three levels of meaning. In F. R. Palmer, editor, *Grammar and meaning*, pages 90 – 116. Cambridge University Press, Cambridge UK, 1995.

[66] D. Lewis. *Counterfactuals*. Harvard University Press, Cambridge MA, 1973.

[67] D. Lewis. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85:297 – 315, 1976.

[68] Seymour Lipschutz. *Set Theory and Related Topics (2nd Edition)*. Schaum's Outline Series, McGraw-Hill, New York NY, 1998.

[69] W. G. Lycan. *Real Conditionals*. Oxford, Oxford, England, 2001.

[70] B. Mates. *Elementary Logic*. Oxford University Press, New York NY, 1965.

[71] M. McDermott. On the Truth Conditions of Certain "If"-sentences. *Philosophical Review*, 105:1 – 38, 1996.

[72] V. McGee. A Counterexample to Modus Ponens. *Journal of Philosophy*, 82:462–471, 1985.

[73] V. McGee. Conditional probabilities and compounds of conditionals. *Philosophical Review*, 98:485–541, 1989.

[74] E. Mendelson. *Introduction to Mathematical Logic (4th Edition)*. Lewis Publishers, New York, 1997.

[75] R. Neapolitan. *Probabilistic Reasoning in Expert Systems*. John Wiley & Sons, New York, 1990.

[76] R. S. Nickerson. *Reflections on reasoning*. Erlbaum, Hillsdale NJ, 1986.

[77] N. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28(1):71–87, 1986.

[78] Donald Nute and Charles B. Cross. Conditional logic. In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic (2nd Edition)*, volume 4, pages 1 – 98. Kluwer, Dordrecht, 2001.

[79] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo CA, 1988.

[80] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York NY, 2000.

[81] W. V. O. Quine. *Word and Object*. MIT Press, Cambridge MA, 1960.

[82] W. V. O. Quine. *Methods of Logic (4th Edition)*. Harvard University Press, Cambridge MA, 1982.

[83] M. D. Resnik. *Choices: An Introduction to Decision Theory*. University of Minnesota Press, Minneapolis, MN, 1987.

[84] K. H. Rosen. *Discrete Mathematics and Its Applications (Fifth Edition)*. McGraw-Hill, Boston MA, 2003.

[85] S. Ross. *A First Course in Probability (6th Edition)*. Prentice-Hall, Englewood Cliffs NJ, 2001.

[86] Bertrand Russell. *A History of Western Philosophy*. Simon and Schuster, New York NY, 1945.

[87] Bertrand Russell. *The autobiography of Bertrand Russell*. Routledge, London, 2000.

[88] Mark Sainsbury. *Logical Forms: An Introduction to Philosophical Logic*. Blackwell, Oxford, England, 1991.

[89] David H. Sanford. *If P, Then Q: Conditionals and the Foundations of Reasoning*. Routledge, London, 1989.

[90] Uwe Schöning. *Logic for Computer Scientists*. Birkhäuser, Boston MA, 1989.

[91] B. Skyrms. *Choice & Chance (4th Edition)*. Wadsworth, Belmont CA, 1999.

[92] S. Soames. *Understanding Truth*. Oxford University Press, Oxford UK, 1998.

[93] R. Stalnaker. *Inquiry*. MIT Press, Cambridge MA, 1984.

[94] Robert C. Stalnaker. Probability and Conditionals. *Philosophy of Science*, 37:64 – 80, 1970.

[95] Robert C. Stalnaker. Indicative Conditionals. In W. L. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, pages 193 – 210. D. Reidel, Boston MA, 1981.

[96] C. L. Stevenson. If-iculties. *Philosophy of Science*, 37:27–49, 1970.

[97] N. I. Stiazhkin. *History of Mathematical Logic from Leibniz to Peano*. MIT Press, Cambridge MA, 1969.

[98] P. F. Strawson. Referring. *Mind*, 59:320–344, 1950.

[99] P. F. Strawson. *Introduction to Logical Theory*. Wiley, New York NY, 1952.

[100] P. Suppes. *Introduction to Logic*. D. Van Nostrand, Princeton NJ, 1957.

[101] E. E. Sweetser. *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge University Press, New York NY, 1990.

[102] Alfred Tarski. *Introduction to Logic and to the Methodology of Deductive Sciences*. Oxford University Press, Oxford, England, 1941.

[103] Katya Tentori, Nicolao Bonini, and Daniel Osherson. Conjunction and the conjunction fallacy. *Cognitive Science*, 2004.

[104] Barbara W. Tuchman. *The March of Folly: From Troy to Vietnam*. Alfred A. Knopf, New York NY, 1984.

[105] A. Urquhart. The complexity of propositional proofs. *The Bulletin of Symbolic Logic*, 1(4):425–467, 1995.

[106] S. Weisler and S. Milekic. *Theory of Language*. MIT Press, Cambridge MA, 2000.

[107] L. Wittgenstein. *Tractatus Logico-Philosophicus*. Routledge & Kegan Paul, London, 1961. German edition published in 1921.

[108] M. Woods. *Conditionals*. Clarendon Press, Oxford UK, 1997.

Index

Index

$\Rightarrow$, 199
$\models_u$, 276, 280
$\models$, 57, 80, 86
$\not\models_u$, 280
$\not\models$, 80
$\not\models \gamma$, 158
$\not\vdash \gamma$, 158
$\varphi\alpha$, 100
$\vdash$, 158
$\vdash_u$, 282
-and-, 200
-or-, 200
TrAs, 50
$\circ$, 114
$\bullet$, 114
$\boldsymbol{Pr}_u$, 285
F, 48
$\mathcal{L}(M)$, 105
$[\varphi]^f$, 284
Meanings, 66
$[\varphi]^t$, 284
$[\Gamma]$, 79
$[\varphi]$, 70
Logic$_u$, 273
NTV, 250, 274
$\boldsymbol{pow}(S)$, 21
$\Rightarrow$, 251
T, 48
U, 262
Reality, 68, 80

a priori justification, 268
Abarbanel-Vinov, Y., 67
abbreviation, 43
abductive logic, 6
absolute events, 231
absolute probability, 231
acceptance, 7
Adams, D., 146
Adams, E., 190, 216, 217, 249, 250, 291, 311
alternative notation, 44
alternative truth tables, 266
ambiguity in natural language, 10
Anderson, A. R., 192
archetypes, 5
argument, 78
Arnauld, A., 4
arrow, 42
artificial language, 8
assertibility, 283, 286
assumption, 114
assumption rule, 115, 153
assumption set, 159
atomic formula, 39

Barbara Bush, 291
Baron, J., 4
basic assignments, 265
basis step, 27
Bayes' Theorem, 243
Beall, J. C., 249, 266, 270, 312
Beaver, D. I., 275
belief, 222
Bennett, J., 194, 249
bets, 226
biconditional, 40
biconditional elimination rule, 139, 156
biconditional introduction rule, 139, 156
Bill Clinton, 84
Bonini, N., 239
Boston marathon, 84
bottom up approach to meaning, 72
Bradley, R., 233, 256, 257
Britney Spears, 298
Buddhism, 34
Bush, George, 275

cancelled assumption, 114
Carson, R., 190
Cartesian product, 24
Castillo, E., 246
category mistakes, 270
chain of reasoning, 109
Changing probability distributions, 234
Chierchia, G., 11, 12, 48
Chihara, C., 228
Chomsky, N., 31, 32
closed under negation, 95
cognitive impenetrability, 270
Cohen, L. J., 231, 255
coherent set of probability claims, 245
collection, 22
collective, 13
collective unconscious, 5
comparing meanings for strength, 69
complement, 18
completeness theorem, 159
compositionality, 56
conclusion, 78
conditional, 40
conditional elimination rule, 117, 154
conditional events, 230
conditional introduction rule, 117, 153
conditional probability, 231, 241
conditional sentences, 5
conditionalization, 235
conditionals, 190
conjunct, 40
conjunction, 40, 51, 101, 266
conjunction elimination rule, 122, 123, 154
conjunction introduction rule, 122, 154
connectives, 34
construction table, 38, 53
contingent formula, 94
contingent meanings, 69
contradiction, 92, 137, 156
contradiction principle, 202
contraposition, 136, 291, 295
Cooper, L., 2

counterfactual conditionals, 193
criterion of adequacy, 197, 288
Cross, C., 194, 250, 312

Dancygier, B., 250
Dawes, R., 4
de Finetti, B., 249, 266, 312
decision theory, 6
declarative sentences, 32
Deduction principle for English, 201
Deduction Theorem, 88, 278, 280
deductive logic, 6, 222
degree of confidence, 222
degrees of belief, 7
DeMorgan laws, 153
DeMorgan rule, 134, 155
DeMorgan, A., 134, 203
derivability of formulas, 149
derivable, 115
derivation, 115, 282
derivation of an argument, 114
derivation schema, 150
derived contradiction rule, 137
derived double negation rule, 132
derived negation elimination rule, 135, 136, 138, 155, 156
derived negation introduction rule, 138
derived rule, 125
descriptive theory, 4
determinate truth value, 33
dinosaur example, 300
disjoint sets, 20
disjunction, 40, 51, 101, 266
disjunction elimination rule, 127, 155
disjunction introduction rule, 127, 154, 155
disjunctive normal form, 102
dissective, 13
distribution, 225, 236
Distribution Law, 153
DNF, 102, 173
dogmatic distribution, 226, 237
domain, 15
double arrow, 42
double negation elimination, 131
double negation in English, 203
double negation rule, 132, 155
double turnstile, 158
Dudman, V., 312
dyadic probability, 231

Ebbinghaus, H. D., 105
Edgington, D., 250, 307
effability, 105
element, 15
Elijah Lagat, 84
empirical justification, 268
empty set, 20
Enderton, H., 188
equivalence classes, 23
Eskimo example, 218, 294
event, 223, 237
exclusive disjunction, 101
exclusive reading of disjunction, 52
expansive, 13
expectation of a bet, 227
expressing conditional probability, 257
expressing meanings, 99
extended probability function, 238

extending a distribution, 242
extension of a truth-assignment, 50

fair bets, 226
falsity conditions, 261, 274, 284
family, 22
finest partition, 23
firmness of belief, 222
first coordinate, 24
first order logic, 188
Flum, J., 105
Fodor, J. A., 270
follows logically from, 81
formulas of $\mathcal{L}$, 35
Freiden, R., 31
function, 25, 26
function defined, 26
function undefined, 26

Gabbay, D. M., 189, 213
Galileo, 2
Gilovich, T., 4
Glymour, C., 4
God example, 219, 294
Goodman, N., 10, 12, 270
grammaticality, 31
Greek letters, 35, 44
Griffin, D., 4
grossest partition, 23
Guenthner, F., 189, 213
Gustason, W., 226
Gutiérrez, J. M., 246

Hacking, I., 7, 226
Hadi, A. S., 246
Hailperin, T., 235
Halpern, J., 67, 235
Harper, R., 67
Harper, W. L., 194, 217, 312
Harry Potter, 191
Hastie, R., 4
Hintikka, J., 56
Hodges, W., 189
Horwich, P., 216
Houston, W., 2
Howson, C., 228

if-then-, 201
Immerman, N., 67
imperative sentences, 32
implication, 106
implies, 43, 80, 88, 160
improper subset, 17
inclusive disjunction, 101
inclusive reading of disjunction, 52
incoherent set of probability claims, 245
independence, 246
independence from context, 48
index, 34
indicative conditionals, 193
indirect proof, 144
induction hypothesis, 27
induction step, 27
inductive logic, 6, 220, 222, 247
ineffability, 105
ineffable meanings, 47
informativeness, 224
interderivable, 150, 169
interrogative sentences, 32
intersection, 18

intransitivity of independence, 246
invalid, 80
invalidating truth-assignment, 83, 111, 183, 279, 282
invalidity in Logic$_u$, 280
Italian, 192

Jackson, F., 215, 250, 287
Jeffrey, R., 6, 109, 188, 235
John, E., 12
Johnson-Laird, P. N., 271, 301
Jung, C., 5

Kahneman, D., 4, 228
Kalmár, L., 159
Katz, J. J., 9, 12, 48
Kay, P., 47
Kempton, W., 47
Kidd, Jason, 191
King Solomon, 88
Kirkham, R., 6
Kleene, S., 190, 266
Kolaitis, P., 67

language of sentential logic, $\mathcal{L}$, 30
Larson, R., 11, 12, 48, 56
Lasnik, H., 31
law of total probability, 240
left hand side, 40, 190
left side strengthening, 217
Levinson, S., 190
Lewis, D., 233, 256, 257, 301
lexical, 10
line, 114
Lipschutz, S., 15
logical consequence, 81
logical equivalence, 97, 282
logical equivalence in Logic$_u$, 283
logical equivalent and assertibility, 286
logical independence, 215
logical independence of variables, 64
long conjunction, 76, 101
long disjunction, 76, 101
love, 2
Lycan, W., 194–196, 249, 250, 253, 259, 289, 297, 311, 312

Madison Avenue, 289
makes true, 57
mapping, 25, 49
marks, 114
Mates, B., 190
mathematical induction, 27, 160, 170, 274
Matsui, Hideki, 197
McConnell-Ginet, S., 11, 12, 48
McDermott, M., 191, 249, 261, 266, 272, 283, 297, 306, 312
McGee, V., 296, 297, 311
meaning, 66
meaning of "or", 187
member, 15
Mendelson, E., 159, 207
Milekic, S., 31
Modus Ponens, 81, 296, 301
Modus Tollens, 81, 295
monadic probability, 231
monotonicity, 217, 279, 293, 308
most important strategic principle, 146

Neapolitan, R., 226
negation, 40, 51
negation elimination rule, 131, 155
negation introduction, 135, 136, 155, 156
negation introduction rule, 130, 155
negation of contradictions, 95
negation of tautologies, 95
nice formulas, 101
Nickerson, R., 9
Nicole, P., 4
Nilsson, N., 235
No truth value thesis, 250
nonatomic formulas, 39
normal disjunctive form, 173
normative theory, 4
not-, 200
number of events, 224
number of meanings, 67
number of truth-assignments, 50
Nute, D., 194, 250, 312

ordered pair, 24
Osherson, D., 239
outcome, 223

partially invalidating truth-assignment, 279
partition, 22, 59, 106
Pearce, G., 194, 312
Pearl, J., 243, 246
personal probability, 226
Pierce's Law, 146, 150
Pierce, C., 146, 149
possible world, 64
power set of a set, 21
pragmatics, 190, 211
precedence, 44
predicate logic, 188
premise, 78
presupposition, 274
Prime minister example, 292, 295
principal connective, 39, 171
principal subformula, 56
principal subformulas of a formula, 41, 170
probability, 222
probability claim, 244
probability distribution, 225
probability distribution for $\mathcal{L}$, 236
probability intuition, 239
probability of conditionals, 253
Probability of events, 229
proper subformula, 37, 41
proper subset, 17
psych experiment on conditionals, 271, 301

Queen Elizabeth example, 216, 269, 293
Quine, W. V. O., 194, 250, 261, 312
quotation, 204

Reagan, Ronald, 220
recursive definition, 35, 50
reductio ad absurdum, 144
refrigerators, 289
Rehoboam, 3
reiteration, 125
Replacement Theorem, 169

representation theorem, 241
representing a distribution, 240
Republican example, 297
Resnik, M., 6, 235, 246
right hand side, 40, 190
Ronald Reagan argument, 297
Rosen, K. H., 15
Ross, S., 222, 230
Rowling, J. K., 191
rule of derivation, 109
Russell, B., 3, 186

Sainsbury, M., 8, 189, 216
sample space, 223, 235
Sanford, D., 93, 215, 217, 218
Santa Claus, 219
satisfiable, 92
satisfies, 57
Schöning, U., 109
second coordinate, 24
secure inference, 6, 83, 199, 222, 268
Segal, G., 11, 12, 48, 56
semantics, 47
sentences of natural language, 30
Sentential Logic, 7
sentential variable, 33
set, 15
set without any members, 19
set-builder notation, 16
simple conjunction, 101, 173
singleton sets, 15
Skyrms, B., 7, 228
Soames, S., 6, 64
sound argument, 84
soundness theorem, 158
Stalnaker, R., 194, 207, 254, 301, 311, 312
standard Sentential Logic, 273
Star Wars, 88
Stevenson, C., 219
Stiazhkin, N., 11
strategic principle, 146
Strawson, P., 190, 275
strength of $\models_u$, 281
strongest meaning, 69
subformulas, 37, 39
subjective probability, 226
subjunctive conditionals, 193
subset, 17
superficial form of sentences, 8
supervaluationism, 249, 266
Suppes, P., 190
Sweetser, E. E., 190
symmetric, 97
syntax, 30

Tagart, J., 271, 301
target argument, 268
Tarski, A., 10
tautology, 86
tautology in Logic$_u$, 276, 281
temporary assumptions, 112
Tentori, K., 239
the meaning of φ, 71
Thomas, W., 105
tilde, 42
top down approach to meaning, 73
torch example, 217, 293
torment, existential, 220
Torre, Joe, 197

total probability, 240
transitive relation, 150
transitivity, 200, 270, 279, 292, 303, 308
transitivity of →, 216
translation scheme, 186
truth conditions, 260, 274, 284
truth functionality, 55
truth functionality of Logic$_u$, 274, 311
truth gaps, 270
truth gluts, 270
truth of a formula, 72
truth preservation, 110
truth table, 59, 60
truth tables, 262
truth value gap, 261
truth values, 48, 264
truth-assignment, 49, 261
truth-functionality, 211
Tuchman, B., 3
turnstile, 158
Tversky, A., 228

unary connective, 57
undefined truth value, 264
uniform distribution, 237
union, 19
universal set, 15
unsatisfiable, 92
Urbach, P., 228
Urquhart, A., 112

vacuous meaning, 68, 86
vague predicates, 11
vagueness, 270
valid, 80
valid in the subset sense, 310
validity in Logic$_u$, 278, 280
van Fraassen, B. C., 249, 266, 270, 312
Vardi, M., 67
variable, 33
vee, 42
Vianu, V., 67
vocabulary of $\mathcal{L}$, 33

weakest meaning, 69
wedge, 42
Weisler, S., 31
Wittgenstein, L., 9
Woods, M., 250