nection between *beliefs*, which may be conclusions arrived at by 'pure reason', and *practical actions* that may be taken as a result of arriving at these conclusions.

Practical Syllogism; Practical Inference A pattern of practical reasoning of the form **want+belief→action**, exemplified in the case of a person who wants candy and who believes that he can buy some in a particular store, who acts accordingly by going to the store.

Pragmatic Conception of Truth Roughly, the view that the truth of beliefs consists in their being useful. Various versions of this doctrine have been advanced by the American philosophers Charles Sanders Peirce, William James, and John Dewey.

Triviality Results Mathematical results proved by David Lewis and generalized by others, that show that a particular measure of the probability of a conditional proposition—that it should be a conditional probability—is inconsistent with the assumption that it satisfies fundamental laws that the probabilities of nonconditional propositions are assumed to satisfy. An implication is that if conditionals' probabilities are measured in this way then they cannot be said to be true or false in senses normally taken for granted in formal logic.

2

Probability and Logic

2.1 Logical Symbolism and Basic Concepts

This chapter will review the concepts of formal symbolic logic that are presupposed in the rest of this text, and introduce basic concepts of probability. Both the concepts and the symbolism of formal logic that we will be using are illustrated in application to an inference that was commented on in chapter 1, which we will eventually consider from the point of view of probability:

Jane will either take ethics or she will take logic. She will not take ethics. Therefore, she will take logic.

Symbolizing the inference in the way that will be used throughout this book, we will write its premises above and its conclusion below a horizontal line, as follows:

$$\frac{E\vee L,\ {\sim}E}{L},$$

or, equivalently, we may write "{E∨L, ~E} ∴ L".[1] The atomic sentences "Jane will take ethics" and "Jane will take logic" are symbolized by the sentential letters E and L, and "either…or…" and "not" (negation and disjunction) are symbolized by "∨" and "~", respectively. "And", "if…then", and "if…and only if…" (conjunction, conditional, and biconditional) will be symbolized likewise, as "&", "→", and "↔", respectively, though chapter 6 will introduce a new symbolism and 'probabilistic interpretation' of the conditional. Later we will use Greek letters for *sentence variables*, for instance when we say that for any sentences ϕ and ψ, ϕ and $\phi{\rightarrow}\psi$ logically

[1]Except in Appendices 7, 8, and 9 we will use a *sentential symbolism* such as "E" for statements like "Jane will take Ethics", and not a *predicate calculus* symbolism like "Ej", where "j" stands for "Jane" and E stands for "will take Ethics". We can do this because, except in the appendices, we will not be dealing with quantified sentences that have to be symbolized with variables.

Henceforth we will usually omit quotation marks around formulas like "E" and "E∨L," and write E and E∨L instead.

entail ψ. This implies that for all particular sentences such as E and L, E and E→L entail L, and the same thing holds for any other formulas that might be substituted for ϕ and ψ.

Formulas that can be formed using the above symbolism are assumed to satisfy the usual logical laws, and the student should be able to construct truth-tables like the one below, to determine the validity of an inference, the logical consistency of its premises, and so on:

	truth values		formulas: state descriptions				other formulas		
cases	E	L	E&L	E&∼L	∼E&L	∼E&∼L	E∨L	∼E	L
1	T	T	T	F	F	F	T	F	T
2	T	F	F	T	F	F	T	F	F
3	F	T	F	F	T	F	T	T	T
4	F	F	F	F	F	T	F	T	F

Table 2.1

This table shows that the premises of {E∨L,∼E} ∴ L are consistent and the inference is valid, since L is true every case in which E∨L and ∼E are both true. In the next section we will see that this implies that if E∨L and ∼E are perfectly certain then L must also be certain, but our problem will be to find out what can be expected if the premises are only probable, and not completely certain. We can't say that the conclusion must be certain,[2] but we can ask: must it at least be probable, and if so, how probable? These questions will be answered in the next chapter, but that will require more background in probability.

We will end this section by introducing a special bit of technical terminology. The formulas E&L, ∼E&L, E&∼L, and ∼E&∼L are sometimes called *state-descriptions* (SDs),[3] and these have two properties that will turn out to be very important when it comes to considering probabilities. (1) Each SD is true in just one 'case', or line in the truth-table. (2) If you know which SD is the true one, you can deduce the truth of any other formula formed just from E and L. For instance, if you know that E&∼L is the true SD, you can deduce that E∨L must be true because E&∼L is only true in line 2 of the table, and E∨L is true in that line.

[2] At least, assuming that it *depends* on the premises. If it is independent of them it must be a logical truth and therefore perfectly certain. But that is not usual in applied logic.

[3] The term 'state-description' is due to R. Carnap (1947), and it is appropriate because a state-description describes the 'state of the entire world', at least so far as concerns propositions formed from E and L.

SDs are also pictured in a special way in *Venn diagrams*, like diagram 2.1 below, which the following section will connect with probabilities:

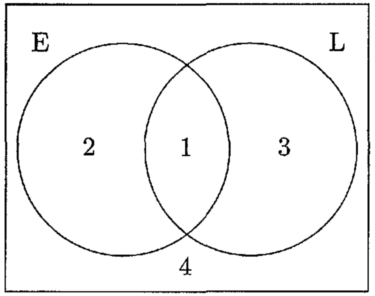


Diagram 2.1

Given a diagram like the above, the student should be able to identify the regions in it that correspond to formulas like E, L, ∼E, E∨L, and so on. For instance, here the *atomic formulas* E and L correspond to circles, and compound formulas like ∼E and E∨L correspond to more irregular regions. ∼E corresponds to everything *outside* of circle E, while E∨L corresponds to the 'union' of circles E and L. More complicated formulas like E→L, (E∨L)↔∼E correspond to more complicated regions, which the student should also be able to identify. This can always be done in terms of unions of the 'atomic *regions*' 1–4, which themselves correspond to SDs.[4] For instance, E→L corresponds to the union of regions 1, 3, and 4 because it is logically equivalent to the disjunction of the SDs E&L, ∼E&L, and ∼E&∼L, which correspond to regions 1, 3, and 4, respectively. The following review exercises will give you practice in doing this.

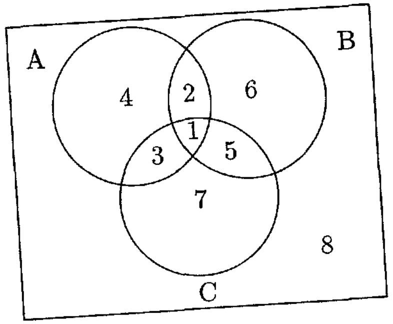
Exercises

1. Give the numbers of the atomic regions in diagram 2.1 that correspond to the following formulas (if a formula doesn't correspond to any regions write "none"):
   a. E↔L
   b. ∼(E&∼L)
   c. ∼(L→E)
   d. E∨(L&∼E)
   e. (E&L)→∼E
   f. (E∨L)↔∼E
   g. (E↔∼E)

[4] Note the difference between atomic *formulas* and the regions they correspond to, and atomic *regions*, which are *parts* of the regions that atomic formulas correspond to. Thus, the atomic formula E corresponds to the left-hand circle in diagram 2.1, whose atomic *parts* are regions 1 and 2. It will turn out later that atomic regions stand to probability logic somewhat as atomic formulas stand to standard sentential logic.

h. $(E\vee L)\&(E\rightarrow L)$
i. $(E\rightarrow L)\rightarrow L$
j. $(E\rightarrow L)\rightarrow(L\rightarrow E)$

2. Determine the atomic regions in the following Venn diagram that correspond to the formulas below the diagram:



a. $A\&B$
b. $C\&B\&{\sim}A$
c. $A\leftrightarrow B$
d. $A\rightarrow(B\&C)$
e. $(A\&B)\rightarrow C$
f. $A\rightarrow{\sim}(A\&B)$
g. $(A\vee B)\rightarrow({\sim}B\vee C)$
h. $(A\rightarrow B)\vee(B\rightarrow C)$
i. $(A\leftrightarrow B)\vee(A\leftrightarrow C)$
j. $A\leftrightarrow(B\leftrightarrow C)$

2.2 Fundamental Connections between Logic and Probability

In order to apply probability to the analysis of inferences like $\{E\vee L,{\sim}E\}$ $\therefore$ L we will add a geometrical interpretation to diagrams like diagram 2.1. This is illustrated in the following three diagrams, the first of which is simply a copy of diagram 2.1:

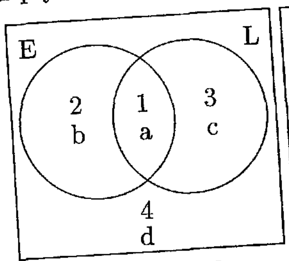


Diagram 2.2

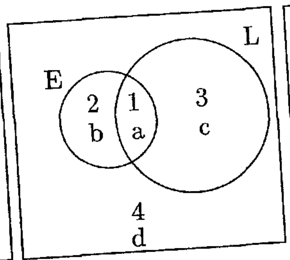


Diagram 2.3

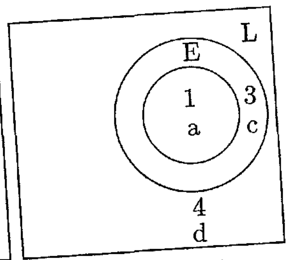


Diagram 2.4

Now we are going to assume that the *areas* of the regions in the diagrams represent the probabilities of the formulas they correspond to (assuming that the area of the 'universe' equals 1). The reason for having more than one diagram is that although all of the diagrams represent the same formulas, the probabilities of these formulas differ in the different diagrams. For instance, while E and L correspond to circles in all three diagrams, and L's probability is actually the same in all three, E's probability is larger in diagram 2.2 than it is in diagrams 2.3 and 2.4. These *possible probabilities* are in some ways similar to and in some ways dissimilar to the possible truth-values that E and L might have.

An extremely important similarity is that the diagrams represent not only the probabilities of E and L but the probabilities of the compound formulas that can be formed from them. For instance, the probability of ${\sim}E$ is represented by the area of the region outside of circle E and the probability of $E\vee L$ is the area of the union of regions E and L. But there is a fundamental dissimilarity between probability and truth.

Note that while E and L have the same probabilities in diagrams 2.3 and 2.4, the probability of $E\&L$ is smaller in 2.3 than it is in 2.4. What this shows is that $E\&L$'s probability is not determined by the probabilities of E and L, which is not like truth, since its truth-value is determined by the truth-values of E and L. This is also the case with $E\vee L$, since while E and L have the same probabilities in 2.3 and 2.4, $E\vee L$ is more probable in 2.3 than it is in 2.4. In the terminology of logical theory, while conjunctions and disjunctions like and $E\&L$ and $E\vee L$ are *truth-functional* in the sense that their truth values are *functions* of the truth-values of E and L, they are not *probability functional* because their probabilities are not functions of the probabilities of E and L.[5] This means that we can't do for probabilities what truth-tables do for truth-values: that is, to calculate the probabilities of compound sentences like $E\&L$ and $E\vee L$ from the probabilities of their parts.

But it is equally important that while the probabilities of formulas like $E\&L$ and $E\vee L$ are not functions of the probabilities of the atomic *sentences* they are formed from, they are functions of the probabilities of the atomic regions they correspond to and the state-descriptions that correspond to them. For instance, $E\vee L$ corresponds to the union of atomic regions 1, 2, and 3, whose probabilities are a, b, and c, respectively, hence $E\vee L$'s probability is a+b+c. It is 'determined' by the probabilities of the state-

[5]This means that probabilities are not 'degrees of truth' as in *many-valued logic* (cf. Rosser and Turquette, 1952) or in the now widely discussed subject of *fuzzy logic* (cf. Zadeh, 1965). There are connections between probability and fuzzy logic, but we shall not enter into them here. Note, incidentally, that the negation of E, ${\sim}E$, *is* probability-functional, which follows from the fact that its probability *plus* the probability of E must always add up to 1.

descriptions, which themselves form a *probability distribution*, as illustrated in table 2.2, which illustrates several other points as well.

There are ten important things to note in table 2.2, as follows:

(1) As previously pointed out, each of the atomic regions corresponds to an SD, and to a case or line in the truth-table, which occupies the black-bordered part of the entire table 2.2.[6] The student should be able to determine which SD and line any given atomic region corresponds to.

(2) As just noted, the set of probabilities or areas of all of the atomic regions constitute a probability distribution, in which probabilities adding up to 1 are 'distributed' among the atomic regions. Another way to picture a distribution is to write the probabilities directly into the corresponding regions in a Venn diagram. For instance, distribution #2 in table 2.2 can be pictured by a diagram like the following, in which areas are not drawn to scale because probabilities are already written into them:

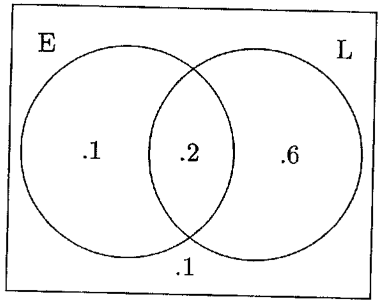


Diagram 2.5

(3) The left-hand columns of table 2.2 list four distributions, namely distributions #1, #2, and #3 plus a 'variable distribution', distribution #4, and the variety of these distributions brings out the fact that the probabilities that form them are entirely arbitrary except for the fact that they must be nonnegative and add up to 1.[7]

(4) Probability distributions are fundamental to probability logic, and possible distributions are to it as possible combinations of truth-values are to standard logic. In particular, they can be used to show that certain combinations of probabilities are *consistent* in the same way that truth-tables can be used to show that certain combinations of statements are consistent. For instance, distribution #2 'generates' probabilities of .9, .7, and .7, for

[6] Usually the converse is true, i.e., every line in the truth-table corresponds to an atomic region, although diagram 2.4 shows that this isn't always so. When a line doesn't correspond to a region we can imagine it corresponding to an 'empty region', whose area is 0.

[7] Assuming that a distribution can be *any* set of nonnegative numbers that add to 1 is what distinguishes probability *logic* from ordinary probability *theory*, which usually restricts itself to special classes of probability distributions.

probability distributions					truth values		formulas							
particular			#4				state descriptions				other formulas			
#1	#2	#3	variable	case	E	L	E&L	~E&L	E&~L	~E&~L	EVL	~E	L	
.25	.1	0	a	1	T	T	T	F	F	F	T	F	T	
.25	.6	1	b	2	T	F	F	F	T	F	T	F	F	
.25	.2	0	c	3	F	T	F	T	F	F	T	T	T	
.25	.1	0	d	4	F	F	F	F	F	T	F	T	F	
							1	2	3	4	1,2,3	2,4	1,2	truth regions
							a	b	c	d	a+b+c	b+d	a+b	dist. #4 prob
							0	1	0	0	1	1	1	dist. #3 prob
							.1	.6	.2	.1	.9	.7	.7	dist. #2 prob
							.25	.25	.25	.25	.75	.5	.5	dist. #1 prob

Table 2.2

E∨L, ~E, and L, respectively, and this proves that it is possible for these propositions to have these probabilities.

(5) However, there are obviously infinitely many possible probability distributions, and distributions #1–#3 are just three of them, while the variable distribution, distribution #4, includes #1–#3 and all other possible distributions as special cases. That there are infinitely many distributions means that you can't list all possible ones in a 'probability distribution table', and use that to determine whether any combination of probabilities is consistent, in the way that you can make a truth-table that lists all combinations of truth values that E and L can have, and use that to determine whether any combination of sentences is consistent.

(6) Distributions #1–#3 have special properties, as follows: distribution #1, which corresponds roughly to diagram 2.2, is a *uniform distribution* since it attaches the same probability to all SDs. It represents 'total uncertainty' as to which of E and L and their combinations is the case, and, as we will see, E and L must both be 50% probable in this distribution.

Distribution #2 is nonuniform, though it still embodies a high degree of uncertainty, and it corresponds roughly to diagram 2.3. It represents a state of affairs in which E is only 30% probable, while L is 70% probable.

Distribution #3, which doesn't correspond to a diagram, is not of a kind that is ordinarily considered in probability theory since it involves only certainties and not probabilities. You could say that ordinary formal logic pretends that these *certainty distributions* are the only possible probability distributions, and the only question is which of them is 'right'.

(7) There is a sense in which distributions #1 and #3 are at opposite ends of a continuum, going from total *un*certainty to complete certainty, with distribution #2 in between.[8]

(8) There is also a sense in which distributions #1–#3 could represent *evolving* probabilities, i.e., distribution #1 could give the 'right' probabilities at one time, distribution #2 could give the right ones at a subsequent time, and distribution #3 could represent the final stage of an inquiry in which all uncertainty is removed. As was noted in chapter 1, the fact that the probability of a proposition can change is another fundamental difference between it and truth. For instance, at one time it can be only 50% probable that Jane will take logic, later it can become 70% probable, and finally it can become certain that she will take it, but the *truth* of "Jane will take logic" remains the same during all that time.[9] How probabilities should change will be returned to, but here the change looks as though it could come about as a result of acquiring *increasing information*, ultimately leading to complete certainty.

[8]It is even possible to measure this 'distributional uncertainty' in terms of Information Theory's concept of *entropy* (cf. section 6 of Chapter 1 of Shannon and Weaver 1949). The entropic uncertainty of the variable distribution is defined as

$$-(a\mathrm{Log}a+b\mathrm{Log}b+c\mathrm{log}c+d\mathrm{log}d),$$

where logarithms are usually taken to the base 2 (the base determines the *unit* of uncertainty measurement). Given this, distribution #1's uncertainty is 2, which is the highest possible uncertainty of a distribution of this kind, distribution #2's uncertainty is 1.626, and Distribution #3's uncertainty is 0, which is the lowest possible uncertainty. Except for footnote comments, *distributional uncertainties* of this kind, which should not be confused with the *propositional uncertainties* that will be discussed in section 2.3, will not be discussed in this book.

(9) The parts of table 1 outside the black-bordered part have to do with how probability distributions determine the probabilities of *formulas*, such as the ones at the heads of the columns on the right. The basic rule is very simple: the probability of any formula formed from E and L is the sum of the probabilities of the cases in which it is true. For instance, E∨L is true in cases 1, 2, and 3, so its probability is the sum of the probabilities of those cases, which are a, b, and c in the variable case. Similarly, because ~E is true in cases 3 and 4 its probability is the sum of the probabilities of those cases, hence it is c+d.

Of course, each SD, E&L, ~E&L, E&~L, and ~E&~L, is true in exactly one case, and therefore its probability is equal to the probability of the case in which it is true.

(10) Looking ahead to chapter 3, something may be said about the light that table 2.2 throws on the inference {E∨L,~E} ∴ L. Distribution #3 makes its premises certain, and, as ordinary formal logic would lead us to expect, its conclusion has become certain in consequence. But distribution #2 may be more realistic, since it makes the premises probable but not certain. It also makes the conclusion probable. We would like to know, however, whether it is necessarily the case that conclusions are probable whenever premises are probable but not certain. This question will be answered in chapter 3.

Exercises

1. Calculate the probabilities of the formulas in exercise 1 of section 2.1, in distributions #1–#4. To simplify, you may want to write out the answers in a table, as follows:

		probabilities in distributions				
	formula	#1	#2	#3	#4	
a.	E↔L	.50	.4	0	a+d	probabilities
b.	~(E&~L)					of formulas
						↓

[9]More exactly, ordinary modern logic considers truth to be 'timeless'. However, in colloquial speech we sometimes speak of something as being true at one time and false at another, and it can be argued that such a view underlies Aristotle's famous 'sea fight tomorrow' argument (cf. McKeon (1941: 48)).

2. a. Can you construct a distribution in which E∨L and ~E have probabilities of .9 and .7, respectively, but the probability of L is less than .7?
   b⋆. What is the smallest probability that L could have, if E∨L and ~E have probabilities of .9 and .7, respectively?
3. Fill in the blanks in table 2.3, which gives possible probability distributions for formulas formed from A, B, and C, corresponding to the regions in Exercise 2 of the previous section.

prob. distributions particular					truth values			formulas			
#1	#2	#3	#4 variable	region	A	B	C	A&B	A↔C	A&(B∨C)	(A→B)∨C
.125	.01	0	a	1	T	T	T				
.125	.02	0	b	2	T	T	F				
.125	.03	0	c	3	T	F	T				
.125	.04	0	d	4	T	F	F				
.125	.1	0	e	5	F	T	T				
.125	.2	0	f	6	F	T	F				
.125	.3	1	g	7	F	F	T				
.125	.3	0	h	8	F	F	F				

A&B	A↔C	A&(B∨C)	(A→B)∨C	
				truth regions
				Dist. #4 prob.
				Dist. #3 prob.
				Dist. #2 prob.
				Dist. #1 prob.

Table 2.3

4. a. Construct a probability distribution for A, B, and C in which all of these formulas have a probability of .9, but A&B&C has a probability less than .8.
   b⋆. What is the smallest probability that A∨B∨C can have if each of A, B, and C have probability .9?
   c⋆. What is the smallest probability that (A&B)∨(B&C)∨(C&A) can have if each of A, B, and C have probability .9?

2.3 Probability Functions and Algebra

Now we will start to use a more compact way of talking about probabilities, and instead of writing "the probability of E∨L", "the probability of ~E", "the probability of L", and so on, we will use *probability function expressions*, "p(E∨L)", "p(~E)", "p(L)", etc.[10] This allows us to express relations involving probabilities much more compactly than before, so that, for instance, instead of saying that the probability of L is less than the probability of E∨L we can write p(L) < p(E∨L). This also allows us to state principles of the *algebra of probability* in a simple way.

As we have seen, the probabilities of compound formulas are not generally *functions* of the probabilities of their parts, but they are in certain special cases. For instance, the probability of ~E is a function of the probability of E, and this can be expressed algebraically by the equation:

$$p(\sim E) = 1 - p(E).$$

More generally, for any formula, ϕ, we can write:

$$p(\sim\phi) = 1 - p(\phi).$$

And, while we can't say that p(L) is always less than p(E∨L) (see diagram 2.4), it is always true that p(L) ≤ p(E∨L), and, more generally, for any formulas ϕ and ψ,

$$p(\phi) \leq p(\phi\vee\psi).$$

The equation and inequality above are particular cases of *general laws of probability* which, taken together, constitute a general *theory of probability*, and, though it is somewhat beside the point for present purposes, it is of some interest to describe some of its features.

⋆2.4 Deductive Theory of Probability

All of the general laws of the theory of probability can be deduced from a very simple set of 'axioms', known as the Kolmogorov Axioms,[11] which state fundamental relations between probability and logic. These are: for all formulas ϕ and ψ

K1. $0 \leq p(\phi) \leq 1$.
K2. If ϕ is logically true then $p(\phi) = 1$.
K3. If ϕ logically implies ψ then $p(\phi) \leq p(\psi)$.
K4. If ϕ and ψ are logically inconsistent then $p(\phi\vee\psi) = p(\phi)+p(\psi)$.

Certain theorems of probability that can be deduced from these axioms will be proved below, but first there are two important things to note

[10] When you read the expression "p(L)" out loud you will still *say* "the probability of L", but, technically, the "p( )" that is common to "p(E∨D)", "p(~E)", and "p(L)" is a *probability function*, and E∨D, ~E, and L are some of its possible arguments.

[11] Cf. Kolmogorov (1950: 2). Kolmogorov's axioms are stated in terms of *sets* rather than formulas and there are other differences of formulation, but for our purposes they are not essential. Sometimes the axioms are stated to include an 'infinite generalization' of K4, which is essential to most advanced mathematical probability theory but not to its present application.