4

Conditional Probabilities and Conditionalization

4.1 Introduction: The Dependence of Probabilities on States of Knowledge

This chapter will turn away temporarily from applications of probability to deductive logic, and consider a 'dimension of probability' that has been left out of account in our discussions so far. This has to do with *conditional probabilities*, which make certain things that probabilities depend on explicit, and the consequences of the fact that what they depend on can change 'in the light of experience'. This sort of change will ultimately have consequences for applications to deductive reasoning because its premises are generally items of information, the learning of which can itself change probabilities. However, except for brief informal comments in this introduction, we will postpone considering the implications for deductive theory to chapter 5.

To illustrate the sort of dependency that we are concerned with, consider the probability of being dealt an ace *on the second card*, when the cards are dealt from a standard, shuffled pack of 52 cards. Assuming that there are 4 aces in the pack, the chance drawing an ace on the second card should be $4/52 = 1/13$. But what if we already know that the first card will be an ace? In these circumstances we would probably say that the chance of drawing an ace on the second card was $3/51 = 1/17$, because this card is 1 of the 51 remaining cards, 3 of which are aces. Thus, the chance of drawing an ace on the second card depends on what is known at the time it is drawn, and when this 'state of knowledge' changes the chances change. Conditional probabilities make this explicit,[1] and if we haven't done so

[1]Conditional probabilities are not usually characterized as quantities that show how 'ordinary', nonconditional probabilities depend on 'knowledge conditions', but this fits in more directly with the applications to deductive reasoning that will be considered in the following chapter. A drawback of this characterization is that it is apt to result in

previously it is because we have tacitly assumed that states of knowledge were unchanging, at least in the contexts that we have been considering.

The fact that conditional probabilities make explicit the changes in probabilities that result from acquiring knowledge makes them especially relevant to logic, and to the theory of knowledge in general, because these subjects are concerned with the acquisition and modification of knowledge by experience and reason. Moreover, there is a law of probability change that is particularly relevant in this connection, roughly as follows: *New experience becomes a part of the state of knowledge that determines probabilities.* The example of the aces illustrates this. Originally, before it is found out that the first card is an ace, the right odds to accept on a bet that the second card would be an ace is the proportion of aces in the whole deck, which is 4/52, but after it is found out that the first card is an ace the right odds to accept on this bet is the proportion of aces in the other 51 cards, which is 3/51. Moreover, the law that relates the *a posteriori* odds of 3/51 to *a priori* odds is simple: 3/51 is the ratio of the *a priori* odds on both the first and second cards being aces to the *a priori* odds on the first card alone being an ace. More generally, the rule is that the *a posteriori* odds on a proposition ϕ, after another proposition, ψ, is learned is the ratio of the *a priori* odds on the conjunction $\phi\&\psi$ to the odds on ψ alone.[2] This is sometimes called Bayes' principle,[3] and, though it is a default assumption that will be examined in section 4.8⋆, we will see that

confusing what people think probabilities are with what they really are. Since 'real' and 'subjective' probabilities are assumed to satisfy the same formal laws, the distinction will not matter until we come to section 9.9⋆⋆, where we discuss the practical value of being 'right' about probability. Footnote 1 of the next chapter also comments briefly on this matter.

[2]In strictness, it should be assumed that the *a priori* probability of ψ is positive, since otherwise the ratio of the *a priori* odds on the conjunction $\phi\&\psi$ to the odds on ψ, which, since it is 0/0, isn't defined. The possibility of learning something that has 0 probability *a priori* cannot be ruled out *a priori*, but obviously it must be very remote, and we will not be much concerned with it. However, the problem of defining ratios of probabilities that are both equal to 0 has been extensively debated in the theoretical literature, and it will be returned to briefly in appendix 2, which concerns V. McGee's generalization of Bayes' principle to the case of new information whose *a priori* probability is 'infinitesimally small'.

Another comment has to do with the terms *a priori* and *a posteriori* (Latin for 'prior' and 'posterior'), which are used here in their standard probabilisitic senses, rather than their philosophical senses. In traditional epistemology (theory of knowledge), *a priori* knowledge was supposed to be something that could be acquired prior to all experience, whereas in probability theory the *a priori* probability of a proposition is simply the probability it had prior to the acquisition of a specific piece of information, which is usually not prior to the acquisition of all information.

[3]Named for Reverend Thomas Bayes (1763) whose posthumously published paper "An Essay Towards Solving a Problem in the Doctrine of Chances" (cf. Bayes, 1940) pioneered the ideas from which the present theory derives. It should be said, however, that the term 'Bayesian', referring to theories of probability and statistics, has no precise

it has enormous implications for logic and the theory of knowledge. In fact, its consequences for inductive reasoning will be discussed at some length in section 4.6 of this chapter, while its consequences for deductive logic will be the subject of the following chapter.

We will begin with a discussion of the formal concept of conditional probability and certain of its properties, before proceeding to consider the full implications of Bayes' principle.

4.2 The Formalism of Conditional Probability

Let A and B symbolize "The first card will be an ace" and "The second card will be an ace", respectively. Then the probability of B, given A, will be written as p(B given A), or more concisely $p(B|A)$, where "|" stands for "given."[4] Thus, we assumed above that while $p(B) = 4/52$, $p(B|A) = 3/51$. Moreover it follows from Bayes' Principle that $p(B|A) = p(B\&A)/p(A)$. But this requires comment.

Although $p(B|A)$ is often defined formally as the ratio $p(B\&A)/p(A)$, in practical applications $p(B\&A)$ would be more likely to be calculated by multiplying $p(B|A)$ and $p(A)$. That is because, while it is obvious intuitively that the chance of getting an ace on the second card given that the first card is an ace is 3 in 51, the chance of getting aces on the first two cards is much less obvious, and the simplest way to calculate it is to multiply $p(B|A)$, which is already known, by $p(A)$. In other words, while $p(B|A) = p(B\&A)/p(A)$ can be regarded as a definition from the formal point of view, in fact $p(B\&A)$ may have to be determined from $p(B|A)$ and not the other way round. For our purposes it is better to think of the equation simply as a rule that allows any one of the three probabilities that it involves to be calculated from the other two.

Generalizing, the conditional probability of any proposition, ϕ, given the knowledge that ψ, will be written as $p(\phi|\psi)$, which is assumed to satisfy the equation

$$p(\phi|\psi) = \frac{p(\phi\&\psi)}{p(\psi)}$$

if $p(\psi) > 0$, and $p(\phi|\psi) = 1$ if $p(\psi) = 0$.[5] For instance, the probability of C,

meaning. What we are here calling 'Bayes' Principle' is a default assumption that can fail in exceptional cases, some of which will be noted in section 4.8⋆.

[4]Sometimes the probability of B given A is written with a slanting stroke, p(B/A), but we want to avoid confusing B/A with a numerical ratio.

"Given" can be understood as meaning "given the knowledge", so that "the probability if getting an ace on the second card, given that an ace has been drawn on the first card" is short for "the probability if getting an ace on the second card, given the knowledge that an ace has been drawn on the first card." This sense of "given" is very close to the one it sometimes has in applied logic, where it might be said that "Jane will take logic" can be inferred, given that she will take either ethics or logic and she won't take ethics.

[5]That $p(\phi|\psi)$ should equal 1 when $p(\psi) = 0$ is a default assumption whose main func-

that an ace will be drawn on the third card, given the knowledge that both the first and second cards are aces, A and B, will be written p(C|A&B), and this is assumed to equal p(C&A&B)/p(A&B). How the unconditional probabilities p(C&A&B) and p(A&B) can themselves be calculated will be returned to in the following section.

Conditional probabilities can also be pictured in Venn diagrams, as *ratios of areas.* For instance, now going back to Jane's classes, the conditional probability of Jane's taking logic, given that she takes ethics, corresponds to the proportion of the 'ethics circle', E, that lies inside the logic circle, L:

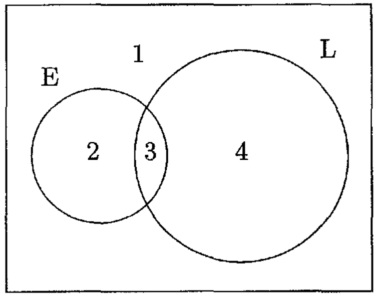

Diagram 4.1

The proportion of E that lies inside L in this diagram is quite small, hence we have represented the probability that Jane will take logic, given that she takes ethics, as being quite small.[6]

There is one more thing to note about diagram 4.1. If you fix attention on the interior of circle E, you see that conditional probabilities *given* E satisfy the same laws as unconditional probabilities. Thus, p(~L|E) = 1−p(L|E) and more generally, if we define the function $p_E(\phi)$ as $p(\phi|E)$ for any proposition ϕ, then p_E satisfies the Kolmogorov axioms.[7]But there

tion is to simplify the theory. However, not all theorists accept this default. Sometimes $p(\phi|\psi)$ is held to be *un*defined when $p(\psi) = 0$, and sometimes it is held to be the ratio of 'infinitesimal probabilities' in this case (cf. appendix 2). This matter is related to the question of whether conditional probabilities are really defined in terms of unconditional ones, since, as noted above, it common for conditional probabilities to be known prior to the unconditional probabilities in terms of which they are supposedly defined. But we will ignore these questions here.

[6]This diagram might suggest that the probability of Jane's tasking logic, given that she takes ethics, could be regarded as the proportion of 'cases' of Jane's taking ethics in which she also takes logic. This would be wrong, however, since Jane can only take these classes once. But there are close connections between conditional probabilities and frequencies of 'cases', some of which are pointed out in section 9.9★★.

[7]More exactly, it satisfies the following modification of the Kolmogorov axioms stated in section 2.3: for all ϕ and ψ,

K1. $0 \leq p_E(\phi) \leq 1$,
K2. if ϕ is a logical consequence of E then $p_E(\phi) = 1$,

is another law of conditional probability that is particularly important in applications.

4.3 The Chain Rule: Probabilistic Dependence and Independence

Recall that we said that Bayes' principle, that $p(\phi|\psi) = p(\phi\&\psi)/p(\psi)$,[8] can be used to calculate any one of the quantities it involves from the other two, and in the case of the cards it is very natural to calculate the probability of getting two aces, p(A&B), as the product p(A)×p(B|A). This rule generalizes as follows:

Theorem 16 (*The chain rule*). If $p(\phi_1\&\ldots\&\phi_{n-1}) > 0$ then $p(\phi_1\&\ldots\&\phi_n) = p(\phi_1)\times p(\phi_n|\phi_1) \times\ldots\times p(\phi_n|\phi_1\&\ldots\&\phi_{n-1})$.

In words: This rule says that the probability of a conjunction of arbitrarily many formulas is the probability of the first conjunct, times the probability of the second conjunct given the first, times the probability of the third conjunct given the first and second, and so on up to the probability of the last conjunct given all of the previous ones. As before, when a theorem involves arbitrarily many formulas it has to be proved by induction, and we will only prove it here in the case in which n = 3.

Theorem 16.3 $p(\phi_1\&\phi_2\&\phi_3) = p(\phi_1)\times p(\phi_2|\phi_1)\times p(\phi_3|\phi_1\&\phi_2)$.

Proof.

1. $p(\phi_2|\phi_1) = p(\phi_1\&\phi_2)/p(\phi_1)$. Definition of conditional probability.[9]
2. $p(\phi_1\&\phi_2) = p(\phi_1)\times p(\phi_2|\phi_1)$. From 1 by pure algebra.
3. $p(\phi_3|\phi_1\&\phi_2) = p(\phi_3\&\phi_1\&\phi_2)/p(\phi_1\&\phi_2)$. Definition of conditional probability.
4. $p(\phi_3\&\phi_1\&\phi_2) = p(\phi_1\&\phi_2)\times p(\phi_3|\phi_1\&\phi_2)$. From 3 by pure algebra.
5. $\phi_3\&\phi_1\&\phi_2$ is logically equivalent to $\phi_1\&\phi_2\&\phi_3$. By pure logic.
6. $p(\phi_3\&\phi_1\&\phi_2) = p(\phi_1\&\phi_2\&\phi_3)$. From 5, by theorem 4.
7. $p(\phi_1\&\phi_2\&\phi_3) = p(\phi_1\&\phi_2)\times p(\phi_3|\phi_1\&\phi_2)$. From 4 and 6.
8. $p(\phi_1\&\phi_2\&\phi_3) = p(\phi_1)\times p(\phi_2|\phi_1)\times p(\phi_3|\phi_1\&\phi_2)$. From 2 and 7. QED

K3. if ψ is a logical consequence of ϕ&E then $p_E(\phi) \leq p_E(\psi)$,
K4. If p(E) > 0 and ϕ&E and ψ&E are logically inconsistent then $p_E(\phi\vee\psi) = p_E(\phi)+p_E(\psi)$.

[8]Assuming as before that $p(\psi) > 0$.

[9]This assumes that $p(\phi_1) > 0$, but that is a logical consequence of the assumption that $p(\phi_1\&\ldots\&\phi_{n-1}) > 0$.