Philosophical Devices:

Proofs, Probabilities, Possibilities, and Sets

David Papineau

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2012

Impression: 2

British Library Cataloguing in Publication Data
Data available

Library of Congress Cataloging in Publication Data
Data available

ISBN 978–0–19–965172–6 (Hbk.)
978–0–19–965173–3 (Pbk.)

Printed in Great Britain by
MPG Books Group, Bodmin King's Lynn

Part III

THE NATURE AND USES OF PROBABILITY

7

Kinds of Probability

7.1 Probabilities of Propositions

Given any proposition p, then we can speak of the probability of p.

For example: the probability that the next card from this pack will be an ace, that this radium atom will decay before the year 3612, that Johnny will go to the party, that it will rain tomorrow, ...

I shall write Pr(p) for the probability of p.

7.2 Kolmogorov's Axioms

In a moment I shall consider what it might mean to say that a certain proposition has a certain probability.

But before that we can note some basic arithmetical constraints. If a way of attaching numbers Pr(p) to propositions p is to count as an ascription of probabilities, it must at least observe the following requirements.

(1) For any p, $0 \leq Pr(p) \leq 1$

(2) If p is certain, $Pr(p) = 1$

(3) If p and q are incompatible, $P(p \text{ or } q) = Pr(p) + Pr(q)$

These are known as Kolmogorov's axioms, and were originally laid out by the great Russian mathematician Andrey Kolmogorov (1903–1987).

The axioms are simple enough. To illustrate, Pr(Johnny goes to the party) is a number between 0 and 1; if it is certain that Johnny will go to the party, then Pr(Johnny goes to the party) = 1; and if Johnny can't go both to the party and the football match, then Pr(Johnny goes to the party *or* the football march) = Pr(Johnny goes to the party) + Pr(Johnny goes to the football match).

7.3 Some Consequences

One immediate consequence of Kolmogorov's axioms is:

(4) Pr(not-p) = 1 – Pr(p)

To see why (4) follows from the axioms, note that p and not-p are incompatible, so by (3)

P(p or not-p) = Pr(p) + Pr(not-p).

But (p or not-p) is certain, so by (2)

Pr(p or not-p) = 1.

The result follows by comparing the right-hand sides of these last two equations.

Here is another useful consequence. In general, whether or not p and q are incompatible:

(5) Pr(p or q) = Pr(p) + Pr(q) – Pr(p and q).

Here 'p or q' should be understood as 'p and/or q', not as 'p or q but not both'. ('Or' will be understood in this sense throughout the book. Logicians call this the 'inclusive' sense, as opposed to the 'exclusive' sense of 'p or q but not both'.)

In this inclusive sense, it will be true that *Johnny goes to the party* or *wears a tie* if he does either on its own and also if he does both, by going to the party in a tie. And so understood Pr(Johnny goes to the party or wears a tie) = Pr(Johnny goes to the party) + Pr(Johnny wears a tie) – Pr(Johnny goes to the party *and* wears a tie).

It is possible to show that (5) follows from Kolmogorov's axioms, but the proof is somewhat laborious, so I shall leave it as an Exercise.

It is much easier to see why (5) must be true by inspecting a Venn diagram. When we look at the diagram, we see that simply adding Pr(p) to Pr(q) would count Pr(p and q) twice—so to get Pr(p or q) we need to correct by subtracting a Pr(p and q). (See Box 15.)

7.4 Joint Probabilities

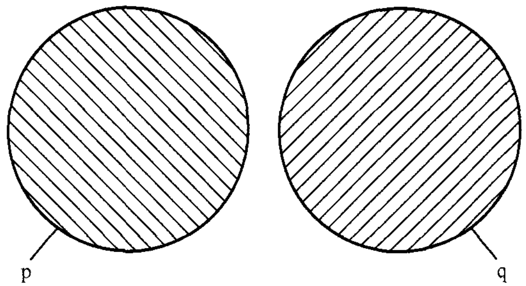
The equivalence (5) told us that

Pr(p or q) = Pr(p) + Pr(q) – Pr(p and q).

However, there is no general rule for the size of Pr(q and p), nor therefore for how much we need to take away from the sum of Pr(p) and Pr(q) to get Pr(p or q). It depends on how much the Venn diagrams for p and q overlap with each other. In our example, it depends on how likely it is that Johnny will both go to the party *and* wear a tie.

We shall consider such joint probabilities—Pr(p and q)—in more detail in the next two chapters, when we discuss conditional probabilities and probabilistic independence. But we can usefully make some initial points here.

In some cases, Pr(p and q) will be zero, namely, when p and q are incompatible—their Venn diagrams don't overlap at all—and then Pr(p or q) will be the simple sum of Pr(p) and Pr(q), as in Kolmogorov's third axiom. This would be the case in our example if there is no way that Johnny would go to the party in a tie.



But in other cases p and q need not be incompatible, and then Pr(p and q) will be a positive number.

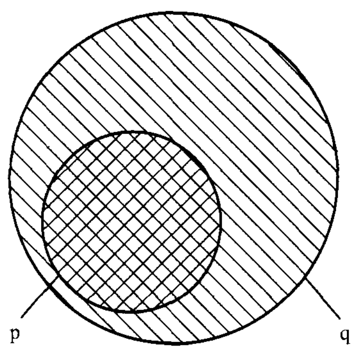
In the extreme case, p will entail q, or q entail p. (For example, Johnny's going to the party may *require* him to wear a tie.)

If p entails q, then the Venn diagram for p is *inside* that for q, so

Pr(p and q) = Pr(p)

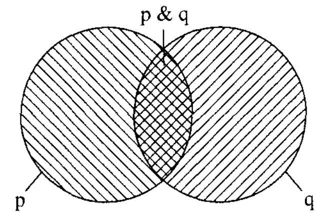
and

Pr(p or q) = Pr(q).



Box 15 Venn Diagrams

This 'Venn diagram' shows why Pr(p or q) = Pr(p) + Pr(q) – Pr(p and q).



In a Venn diagram we take the points in a plane to represent possible worlds, and so can use sets of points to represent sets of possible worlds, and in particular to represent all those possible worlds where some proposition p is true. The areas of these spaces can then be used to represent the probabilities of the relevant propositions. (Note here how it is possible to equate a proposition with the set of possible worlds where it is true. This equivalence is widely used in philosophy.)

In the above diagram the proposition ***p or q*** corresponds to the points which are either in the area labelled p, or in the area labelled q, or in both. And the proposition ***p and q*** corresponds to the points which are in both the area labelled p ***and*** in the area labelled q—that is, the cross-hatched area.

It is easy to see that, if we tried to work out the area corresponding to ***p or q*** by simply adding the area for p to that for q, we would count the cross-hatched area twice. So to get the right answer we need to correct by subtracting the cross-hatched area.

In our example, if Johnny's going to the party requires him to wear a tie, then the Venn diagram for Johnny's going to the party will be inside the one for his wearing a tie, so

Pr(Johnny goes to the party *and* wears a tie) = Pr(Johnny goes to the party)

and therefore

Pr(Johnny goes to the party *or* wears a tie) = Pr(Johnny wears a tie).

If q entails p, then the Venn diagram for q is inside that for p, and these results are reversed.

So Pr(p and q) can sometimes equal Pr(p) and sometime equal Pr(q) (when p entails q or when q entails p respectively).

But note that Pr(p and q) can never *exceed* either of these numbers. Pr(Johnny goes to the party *and* wears a tie) can't be greater than either Pr(Johnny goes to the party) or Pr(Johnny wears a tie).

Sometimes it is easy to forget this. (See Box 16.) But you shouldn't. Two things both happening (p *and* q) can never be more likely than either one happening on its own.

7.5 Subjective and Objective Probabilities

There are two quite different ways of interpreting probability statements—that is, of understanding what it means when we attach numbers between 0 and 1 to propositions in such a way as to satisfy Kolmogorov's axioms of probability.

We can understand such statements either as reports about *subjective* probabilities or as reports about *objective* probabilities.

Subjective probabilities measure the extent to which *agents expect* outcomes. Objective probability measures the *real tendencies* for those outcomes to occur.

Box 16 Linda the Feminist Bank Teller

Let me tell you about Linda. She is 31 years old, single, outspoken, and very bright. She did an undergraduate degree in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Now, which of these propositions is more probable?

(A) Linda is a bank teller.

(B) Linda is a bank teller and is active in the feminist movement.

It is very natural to choose (B). When the psychologists Daniel Kahneman and Amos Tversky tested people on this question, they found that about 9 out of 10 chose (B). Indeed, when they tested doctoral students in the decision science programme at Stanford Business School, a group with an intensive training in probability and statistical theory, they still found that over 8 out of 10 chose (B).

Yet (B) cannot be the right answer. Two things cannot be more likely than one. After all, in every situation where Linda is a bank teller *and* a feminist, she will also be a bank teller, and in addition there will be situations where she is a bank teller without being a feminist.

Something about the Linda story confuses our thinking. (If you're not convinced that (B) is wrong, it might be helpful to think in terms of money. Suppose you are going to win £100 for a correct answer. Would you rather commit yourself to (A) or to (B)?)

7.6 Subjective Probability

Imagine that you are going out for a short walk, and you take both your sunglasses and your umbrella. Do you believe it is going to rain?

Well, you aren't certain it *is* going to rain—otherwise why take your sunglasses?

But you aren't certain that it is *not* going to rain either—otherwise why take your umbrella?

In a case like this, it seems natural to say that you have a certain *degree of belief* in the proposition *it will rain*, and that this can be represented by some number between 0 and 1. (If you were certain it will not rain, then your degree of belief would be 0, and if you were certain it will rain, then your degree of belief would be 1.)

Alternative names for these degrees of belief are 'subjective probabilities' or 'personal probabilities' or 'credences'.

7.7 Action, Utility, and Subjective Probability

We can think of degrees of belief as manifesting themselves in choices of actions (as when you took both your umbrella and your sunglasses in the example above). In general, the greater degree of belief an agent attaches to some proposition p, the more that agent will be inclined to perform actions that will bring good results *if* p.

The easiest way to connect degrees of belief with choice of actions is to focus on *betting* behaviour. Given some proposition p, ask yourself how much you would be prepared to pay for a bet that will pay £1 *if* p. (For example, how much are you prepared to pay to win £1 *if Johnny comes to the party*?) The fraction of £1 that you are prepared to stake plausibly measures your degree of belief in p. You'll be prepared to bet 50p if your degree of belief is 0.5, but only 10p if your degree of belief is 0.1.

Maybe you don't think of yourself as much of a gambler. But note that pretty much any action can be construed as a gamble. When you cross the road, this is presumably because your degree of belief that you will get to the other side (a good result) is very much bigger than your degree of belief that you will be run over (a very bad result).

Many philosophers and economists hold that, in general, when someone performs an action, this is because the *expected utility* of that action is greater than that of the alternative actions available. The idea here is that the agent is concerned about certain outcomes (getting to the other side, being run over) whose importance can be measured by some positive or negative number—its '*utility*'. And the *expected utility* of an action is then the sum of those utilities each multiplied by the agent's degree of belief that the action will lead to that outcome.

Thus suppose the utility of getting to the other side is plus 10, and your degree of belief that crossing the road will lead to this is 0.9999; and the utility of being run over is minus 10,000, and your degree of belief for this is 0.0001. Then the expected utility of crossing the road will be:

$$(10 \times 0.9999) + (-10{,}000 \times 0.0001) = 9.999 - 1 = 8.999$$

and this may well be higher than the expected utility of the alternative actions currently open to you.

Of course all this is at best a kind of idealization. In truth, there isn't really a precise answer to the question of exactly how much I believe p, for every proposition p. There are plenty of propositions that I have never thought of, and even among those I have thought of are many to which I have a pretty fuzzy attitude. Nor is it very realistic to suppose that I can attach numbers to all the things I care about. Still, perhaps we can go along with the idealization in order to simplify the arguments that follow. (Compare the way in which engineers simplify their calculations by assuming that everyday objects like a block of concrete have precise masses, even though in truth it will always be a bit vague whether some of the molecules on the surface are attached to the block or not.)

So I shall assume henceforth that for any person X, at any time t, and any proposition p, there will be a number between 0 and 1 that represents X's degrees of belief at time t in proposition p.

7.8 Dutch Books

I said that degrees of belief or subjective probabilities offer one way of interpreting probability statements—that is, one way of attaching numbers between 0 and 1 to propositions in such a way as to satisfy the axioms of probability.

However, as yet I haven't really shown this, for I haven't yet shown that degrees of belief do satisfy the axioms of probability.

And in fact there is no guarantee that they will. Nothing in psychology rules out the possibility that an agent at a time might attach a degree of belief 0.6 to the proposition *it will rain* and simultaneously a degree of belief 0.6 to the proposition *it won't rain*, thus violating the immediate implication of the probability axioms that Pr(p) = 1 − Pr(not-p). (Maybe the agent wasn't thinking very hard, and somehow managed to take a positive view of both these propositions at the same time.)

However, there is an argument that any *rational* degrees of belief must conform to the axioms of probability, even if *actual* degrees of belief don't always do so.

The argument is that anybody whose degrees of belief violate the axioms of probability can have a *'Dutch Book'* made against them. A Dutch Book is a set of bets which are *guaranteed to win whatever happens.*

By way of illustration, consider the person who believes *it will rain* to degree 0.6 and also believes *it won't rain* to degree 0.6. Well, this person will happily pay 60p to win £1 on its raining, and also happily pay 60p to win £1 on its *not* raining. But anybody who makes this pair of bets will certainly lose whatever happens, because they will have paid out £1.20 in total and will only win £1 whether it rains or not.

It is not hard to prove that a Dutch Book can be made against you if and only if your degrees of belief fail to satisfy the axioms of probability.

(The subject in the above illustration got into trouble because of degrees of belief in p and not-p which added to more than 1. This might

make it seem safe to have degrees of belief that add to *less* than 1. However, in that case you could be induced to bet *against* both p and not-p in a way that is guaranteed to lose.)

Since it seems clearly irrational to adopt attitudes that can make it certain that you will incur a loss, it follows that any rational agent will have degrees of belief that do conform to the probability calculus. (Such agents are called 'coherent'; those whose degrees of belief violate the axioms are 'incoherent'.) (See Box 17.)

Note that there is nothing in this 'Dutch Book Argument' to specify *what* degrees of belief you should have, beyond requiring that they must conform to the probability axioms. You can be coherent by having a subjective probability of 0.6 for *it will rain* and of 0.4 for *it won't rain*. But you could equally achieve coherence by attaching 0.8 and 0.2 to these two propositions, or 0.15 and 0.85, or any other combination of numbers that add up to 1.

The 'Dutch Book Argument' requires coherence, but beyond that leaves it to subjective opinion which particular degrees of belief you should adopt.

7.9 Objective Probability

Objective probabilities are quite different from subjective ones. They are out in the world, not in people's heads. They quantify the objective tendencies for certain kinds of results to happen. These tendencies would still have existed even if agents with subjective probabilities had never evolved.

The clearest examples of objective probabilities come from the quantum mechanics of subatomic processes. Certain events at this level are absolutely unpredictable. Take any radium atom. It may decay in a given time interval or it may not. There is no difference between those atoms that decay and those that don't. All that can be said is that each such atom has a certain objective probability of decaying in a given interval. (If the interval is 1602 years—the 'half-life' of a radium atom—then there is 0.5 probability of decay in that time.)

It is helpful to think of objective probabilities in terms of *frequencies*. If the probability of a single radium atom decaying within its half-life is 0.5, then about 50% of any sequence of radium atoms will decay in that time.

(But don't be too quick to *equate* objective probabilities with frequencies. There are many philosophical pitfalls in the way of any such equation, most centrally the fact that the observed frequency in any sequence of events won't generally correspond *exactly* to the underlying probability. Note how I was careful to say above that '*about* 50% of any sequence of radium atoms will decay in that time'—not that exactly 50% will.)

There are plenty of objective probabilities outside the subatomic world (though perhaps they all depend in some way on quantum probabilities). For example, the probability that any human embryo will be male is slightly over 0.5. The probability that males in the United States will develop pancreatic cancer in their lifetime is 0.0138. The probability that an ace will be dealt first from a well-shuffled pack is 1/13. And so on.

The ultimate nature of objective probability is a matter of philosophical controversy. But we need not enter into this here. The basic point is that objective probabilities are genuine features of the external world, distinct from subjective degrees of belief.

Box 17 Bookmakers and Dutch Books

A good bookmaker aims to make a Dutch Book against the punters. The bookie wants to induce the punters to make a set of bets that will turn a profit for the bookie whichever horse wins.

For instance, in a two-horse race between Aramis and Balthazar, the bookie will be guaranteed a profit whichever horse wins if £100 has been staked on Aramis at evens, and £120 on Balthazar at 2-1 on. ('Evens' means that you stake £1 to win £1, and '2-1 on' means you stake £2 to win £1.) These bets mean that the bookie will make £20 if Aramis wins (the £120 stake on Balthzar less the £100 payout on Aramis) and £40 if Balthazar wins (the £100 stake on Aramis less the £60 paid out on Balthazar).

This doesn't necessarily mean that any individual punter is irrational. The bookie can pull this trick because different punters will sometimes attach different subjective probabilities to the same outcome. In this sense the punters *taken collectively* will violate the axioms of probability. But this doesn't mean that any individual punter has 'incoherent' degrees of belief.

But you will be irrational if the bookie can make a Dutch Book against you all *on your own*. If you yourself put £100 on Aramis at evens, and *also* put £120 on Balthasar at 2-1 on, then this indicates that you personally have a degree of belief in Aramis winning of at least 1/2 and in Aramis *not* winning of at least 2/3. Now the bookie is not only sure to win, but you individually are sure to lose.

FURTHER READING

Two of my old teachers have written excellent philosophical introductions to probability:

An Introduction to Probability and Inductive Logic by Ian Hacking (Cambridge University Press 2001).

Probability: A Philosophical Introduction by D. H. Mellor (Routledge 2005).

The Stanford Encyclopedia entry by Alan Hayek is a thorough discussion of the different 'interpretations of probability': <http://plato.stanford.edu/entries/probability-interpret>.

Daniel Kahneman's *Thinking, Fast and Slow* (Allen Lane 2011) explains how humans are very prone to mistakes in probabilistic reasoning.

EXERCISES

1. If I draw one card from a well-shuffled pack, what is the probability of:

   (a) a heart
   (b) a king
   (c) an honour (A, K, Q, J, 10)
   (d) not a heart
   (e) an honour and a heart
   (f) a heart or a spade
   (g) a heart and a spade?

2. If I toss a fair coin four times, what is the probability that I get:

   (a) four heads; (b) zero heads; (c) one head; (d) three heads?

   Hint: there are 16 equiprobable outcomes for the four-toss sequence.

3. If I roll two fair dice, what is the probability that they sum to:

   (a) 4; (b) 7; (c) 12; (d) an odd number; (e) less than 5; (f) either less than 5 or 9; (g) either less than 5 or an even number?

   Hint: there are 36 equiprobable ways the dice can land.

4. If Pr(Johnny at party) = 0.4 and Pr(Jenny at party) = 0.8 and Pr(Johnny and Jenny at party) = 0.3, what is the probability that

   (a) Jenny won't be there
   (b) at least one of them will be there
   (c) Jenny will be there but not Johnny?

5. Suppose that you can either go to the beach or to watch the test match. The beach has an intrinsic utility of plus 10, and the cricket of plus 15. But there is a 0.5 chance that you will get sunburnt (utility of minus 10) at the beach, where there is only a 0.3 chance of getting sunburnt at the cricket. Also, there is a 0.2 chance you will see Jill (plus 20) at the beach, but only a 0.05 chance you will see her at the cricket. Which option has the greater expected utility?

6*. Show algebraically how the equation

   Pr(p or q) = Pr(p) + Pr(q) – Pr(p and q)

   follows from Kolmogorov's axioms. (Hint: note that

   (p or q) is logically equivalent to ((p & not-q) or (q))

   and that

   p is logically equivalent ((p & q) or (p & not-q))

   and that the pairs of propositions within the brackets on the right-hand sides are incompatible.)

8

Constraints on Credence

8.1 The Principal Principle

The last chapter ended with the contrast between subjective and objective probabilities. Some readers might have wondered how they are related.

Not every proposition to which agents attach subjective degrees of belief will also have an objective probability. You might well have a certain expectation of Johnny going to the party, say, or of Aramis winning the 3.30 at Kempton Park, even if there is no good sense in which these propositions have any objective probability.

But in other cases agents do attach subjective degrees of belief to propositions that also have an objective probability—for example, that a given atom will decay in some interval, or that a given embryo will be male, or that the next card drawn from a well-shuffled pack will be an ace.

Now, there is no guarantee in such cases that the agent's subjective probability will correspond to the objective probability. You might *expect* an ace to degree 1/2, even though its objective probability is only 1/4.

But even so there is something obvious to say about the relation between subjective and objective probability in such cases—namely

The Principal Principle:
An agent's subjective probabilities *ought* to match the objective probabilities, even if in fact they don't.

The term 'Principal Principle' was originally coined by David Lewis (the same philosopher who was a realist about possible worlds) for his version of the idea that subjective probabilities ought to match objective probabilities. He adopted this name because he thought that this idea is fundamental to our understanding of both objective and subjective probability.

In fact my Principal Principle above is only a rough approximation to Lewis' more carefully formulated principle. But it will do for present purposes.

Remember that the 'Dutch Book Argument' allowed rational agents a great deal of freedom about the choice of subjective probabilities—the only constraint was that subjective probabilities should conform to the axioms of probability. The Principal Principle imposes a further constraint on rational agents—when objective probabilities exist, you should do what you can to make your subjective probabilities match them.

The Principal Principle is obviously sensible. If you are to make the right choices, your subjective expectations had better not diverge from the objective probabilities. You will make bad bets if you have a high degree of belief that an ace will be dealt, when in fact the objective probability is only 1/4.

Curiously, even though conformity to the Principal Principle is obviously a good idea, the status of this principle is a matter of controversy. Some philosophers think it can be justified by appeal to more basic facts. But others doubt that any such justification is possible, and view it as itself a fundamental principle of rationality.

8.2 Conditional Probability

The *conditional probability* of p given q, Pr(p/q), is the probability to ascribe to p on the assumption that q.

It is measured by:

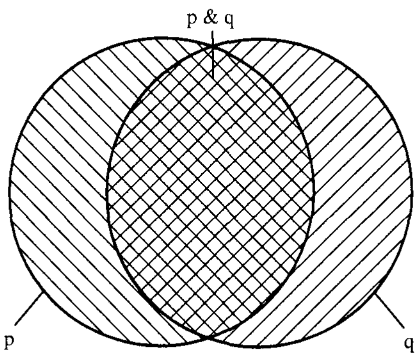
(1) Pr(p/q) = Pr(p & q)/Pr(q).

(In Venn diagram terms, think: the area of q that is also p—that is, the cross-hatched area as a proportion of the area for q.)

So, for example, we might have the conditional probability that a throw of a fair die will show an even number, given that it shows a higher number than three. We can write this Pr(even/over three), and measure it by:

Pr(even *and* over three)/ Pr(over three).

This fraction represents the probability of an even result among the results that are higher than three—and is equal to 2/3, since the probability of a result (four or six) that is even *and* over three is 2/6, while the probability of *any* result over three (four, five, or six) is 1/2.



8.3 Updating Degrees of Belief—Conditionalization

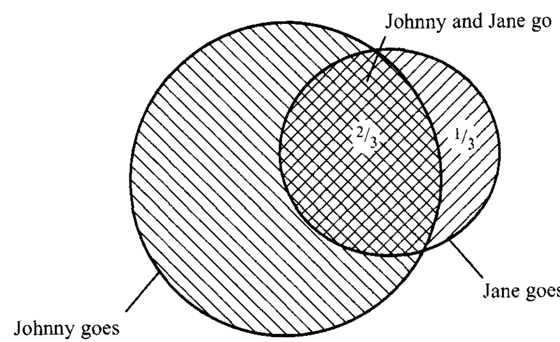
Now that we have introduced conditional probabilities, we can explain a further constraint governing rational degrees of belief. So far we have seen how the 'Dutch Book Argument' implies that rational degrees of belief must be *coherent* (that is, satisfy the axioms of probability), and how the Principal Principle implies that they must *match objective probabilities* when these are available. The further constraint is that rational agents should *'conditionalize'* whenever they gain new information.

Suppose that you have rational degrees of belief as follows:

Pr(Johnny goes to the party) = 1/2
Pr(Johnny goes to the party/Jane goes to the party) = 2/3.

Now you learn for sure that Jane is going to the party. What should your degree of belief in Johnny's going now be?

The answer is obvious enough—2/3. If it was right to think beforehand that the conditional probability of *Johnny going/on the assumption Jane goes* is 2/3, and if now it turns out that Jane *is* going, then it must be



right to think that the unconditional probability of Johnny has increased to 2/3.

Think of it in Venn diagram terms. You now know you are *inside* the area of the Venn diagram for Jane's going, so to speak. And you have already decided that the proportion of this area that covers Johnny going is 2/3. So it must now be rational for you to have an unconditional degree of belief in Johnny going of 2/3.

Changing your degrees of belief in this way is called '*conditionalization*'. Let us thus formulate

The Principle of Conditionalization:
If your *old* conditional degree of belief $Pr_{old}(p/q)$ equals k, and you come to know q, you should set your *new* degree of belief in p, $Pr_{new}(p)$, equal to k.

Note that q here needs to be understood as representing *everything* you come to know. The principle doesn't work if q is only *part* of your new knowledge.

Thus suppose that in the above example you learn not only that Jane is going to the party but also that she will be accompanied by Jill. And suppose that you had always thought that there was almost no chance that Johnny would go if *both* Jane and Jill did. (You had a very low original conditional probability Pr_{old}(Johnny goes/Jane *and* Jill go) even though your original Pr_{old}(Johnny goes/Jane goes) was 2/3.)

While it is still true that you have learned that *Jane will go*, it is no longer a good idea to attach a 2/3 probability to Johnny going, just on the grounds that your Pr_{old}(Johnny goes/Jane goes) = 2/3. And this is precisely because you have learned *more* than that Jane will go to the party. You now know not just that you are inside Jane's Venn diagram, so to speak, but more specifically that you are inside that bit of it where Jill also goes to the party. And the proportion of *that* area where Johnny goes too is very small.

It is generally agreed *that* the Principle of Conditionalization is valid. But, just as with the Principal Principle, there is no agreement about

why it is valid. As before, some philosophers think it is a basic principle of rationality, while others think that it can be justified by further considerations.

Note in this connection that the Principle of Conditionalization is *not* simply a consequence of the Dutch Book Argument for coherence. That earlier argument showed that the axioms of probability must be respected by all the degrees of belief you adopt *at any given time*. But the Principle of Conditionalization concerns the way you should *change* your degrees of belief *over time* in response to evidence, substituting your old degrees of belief Pr_{old}(—) by new ones Pr_{new}(—).

You will satisfy the Dutch Book Argument as long as your old Pr_{old}(—)s and your new Pr_{new}(—)s are each separately coherent. The Principle of Conditionalization places a further constraint on the how these two sets of degrees of belief are related.

8.4 Bayes' Theorem

There is a simple probability equation that casts some useful light on the workings of conditionalization:

(2) $Pr(h/e) = Pr(h) \times Pr(e/h)/Pr(e)$.

This equation, which you can check follows very quickly from the equation (1) for conditional probability, is known as *Bayes' Theorem*, after the eighteenth-century English clergyman who first proved it.

To see the significance of this equation, consider some case where you gain some evidence e and are concerned with its bearing on some hypothesis h. According to the Principle of Conditionalization, you should adopt a new $Pr_{new}(h)$ that is equal to your old conditional $Pr_{old}(h/e)$. But Bayes' Theorem tells us that $Pr_{old}(h/e)$ is equal to $Pr_{old}(h) \times Pr_{old}(e/h)/Pr_{old}(e)$. So we can see that the two together imply that

(3) $Pr_{new}(h) = Pr_{old}(h) \times Pr_{old}(e/h)/Pr_{old}(e)$.

We can view this as a recipe for transforming your old degree of belief in h to a new one when your learn e—multiply your $Pr_{old}(h)$ by the factor on the right-hand side. This tells you that you should increase your degree of belief in h to just the extent that $Pr_{old}(e/h)$ exceeded $Pr_{old}(e)$—that is, to just the extent that e was to be expected given h but not to be expected otherwise.

So viewed, (3) seems eminently sensible. The hypothesis h is confirmed if it successfully predicts something that would otherwise be unexpected.

In addition to thus explaining why a hypothesis gains more credibility from the verification of *surprising* rather than unsurprising consequences, Bayes' Theorem also illuminates a wide range of other quirks and puzzles about the way evidence confirms hypotheses.

For example, (3) explains why it is a mistake to ignore the prior probability of h in assessing how probable it is shown to be by e. (This surprisingly common mistake is known as the 'base rate fallacy'. See Box 18.)

Because of the significance of Bayes' theorem, the term *'Bayesian'* is often found in discussions of probability. However, this term has no very definite meaning. It is probably most often used to refer to any view that takes subjective degrees of belief seriously and holds that they are subject to some rational principles. But sometimes it is used more precisely, to refer specifically to the idea that degrees of belief should be updated according to the Principle of Conditionalization.

8.5 Conditional Probabilities and Conditional Statements

A conditional probability Pr(q/p) is the probability of q *on the assumption* that p.

Some readers might have wondered how such conditional probabilities relate to *conditional statements* of the form *if p, then q*. (For example: *if Jane goes to the party, then Johnny will go too.*)

After all, doesn't a conditional statement amount to something like stating q *on the assumption* that p? And given this, shouldn't we expect the probability of the conditional statement Pr(if p, then q) to be equal to the conditional probability Pr(q/p)?

As it happens, this is a horribly complicated topic.

An initial difficulty is that there are different kinds of conditional statement. In a moment I shall distinguish between *material*, *indicative*, and *subjunctive* conditionals. And even after we have distinguished them, it is not obvious how to understand them. While material conditionals are clear enough, the analysis of indicative and subjunctive conditionals is hugely controversial.

It would take us too far afield to analyse these constructions properly here. My aim in the brief remainder of this chapter will simply be to show you why we need to recognize different kinds of conditionals.

What about the question with which I started this section—is the probability of a conditional statement Pr(if p, then q) equal to the conditional probability Pr(q/p)? Here I can do no more than simply tell you that this simple equation doesn't work for *any* kind of conditional 'if..., then' statement—which is not to deny that there are important connections between conditional statements and conditional probabilities.

8.6 Material Conditionals

If you have done an elementary logic course, you will have been introduced to a construction, normally written 'p→q', which is defined as being true as long as it is not the case that p is true and q is false.

This is the 'material conditional'.

Given its definition, it is easy to see that 'p→q' is equivalent to 'not-(p and not-q)' or again to 'either not-p or q'.

It is normal in elementary logic courses to read 'p→q' as 'if p, then q'.

And indeed the material conditional does have strong similarities with everyday claims of the form 'if p, then q'. In particular, it shares the feature that, when you add knowledge of p to them, then you can infer q. Just as p together with 'if p, then q' implies q, so does p together with 'p→q'. (This is an immediate consequence of the definition of 'p→q' given above—you can check it as an exercise.)

Given the similarities, there is no great harm in reading 'p→q' as equivalent to everyday claims of the form 'if p, then q' when exploring elementary logic. But there are strong reasons to doubt that the two constructions are really the same.

Note that 'p→q' is guaranteed to be true whenever p is false, whatever q says, and also to be true whenever q is true, whatever p says. (Remember, 'p→q' is true as long as it is not both the case that p is true and q is false.)

So 'David Papineau goes to Antigua in November → the gold price rises in December' is guaranteed to be true, as long as I do not go to Antigua in November.

Similarly 'Cesc Fabregas plays for Arsenal → Hugh Grant lives in London' is guaranteed to be true, simply in virtue of Hugh Grant living in London.

Now, as we shall see in a moment, the everyday construction 'if…, then…' can be used to make two different kinds of claim—'indicative' and 'subjunctive' conditional claims. But we can already see reasons why the material conditional 'p→q' must differ from both of these. In ordinary English, any claim of the form 'if p, then q' requires some *connection* between p and q, not just the falsity of the antecedent p or the truth of the consequent q.

So, on any reading of the English construction 'if…, then…', my not going to Antigua in November isn't enough to ensure the truth of '*if* David Papineau goes to Antigua in November, *then* the gold price will rise in December'—for there may be no connection between my November location and the December gold price.

Box 18 The Base Rate Fallacy

You are worried about a kind of cancer (h) which is present in 1% of people like you. There is a simple test which invariably detects the cancer, though it does give a false positive result in 10% of people without it. You take the test, and get a positive result (e). What now is the probability you have the cancer?

Well, you might think that, since the test is only 10% unreliable, the answer must be 90%. But that would be quite wrong. There is still little more than a 9% probability of cancer.

To see why, recall that, once you discover e, you should set your new $Pr_{new}(h)$ equal to your old $Pr_{old}(h/e)$. And Bayes' Theorem tells you to compute this by multiplying your old $Pr_{old}(h)$ by $Pr_{old}(e/h)/Pr_{old}(e)$.

Two of these terms are easy. $Pr_{old}(h)$ was given as 1%, and $Pr_{old}(e/h)$ is 1, since the test invariably detects the cancer. $Pr_{old}(e)$ is a bit messier: what is the probability of a positive result for a person taken at random? Well, the 1% of cancer sufferers will definitely give positive results, and the 99% of non-sufferers will give 10% false positives—which sums to 10.9%. So $Pr_{old}(h) \times Pr_{old}(e/h)/Pr_{old}(e) = 0.01 \times 1/0.109 \approx 0.0917$. So you should set your $Pr_{new}(h)$ to just over 9%.

Think of it like this. If 1,000 people take the test, 10 will give a positive result because they have the cancer—but 99 healthy people will give false positives. So a bad result still leaves you with only a $10/109 \approx 0.0917$ probability of cancer.

The tendency to overestimate the significance of such tests is called the 'base rate fallacy', because it is due to ignoring the low 'base rate' or initial probability of having the cancer. It is disturbingly common in everyday life.

And similarly Hugh Grant's living in London isn't enough to ensure the truth of '*if* Cesc Fabregas plays for Arsenal, *then* Hugh Grant lives in London'—for Cesc Fabregas' employment may have nothing to do with Hugh Grant's residence.

Given these differences, it seems clear that the material conditional works differently from any version of the everyday construction 'if p, then q'. (Indeed, we might feel that 'material *conditional*' is something of a misnomer, given its marked difference from any everyday 'if p, then q'.)[1]

8.7 Indicative and Subjunctive Conditionals

Consider this pair of claims.

(4) 'If Oswald didn't kill Kennedy, then someone else did.'

This claim is obviously true. There is no doubt that President Kennedy was killed by somebody. If Lee Harvey Oswald wasn't in fact the guilty party, then some else must have done it.

(5) 'If Oswald hadn't killed Kennedy, then someone else would have.'

This claim is very doubtful. The Warren Commission investigated the matter very thoroughly and concluded that Oswald was working alone. In their view, if Oswald's plans had somehow been frustrated, then Kennedy would not have been killed—that is, they concluded that (5) is false.

Since (4) is clearly true and (5) very likely false, they must mean different things.

[1] I should note that there are a few philosophers who maintain that the indicative version of the everyday 'if p, then q' is at bottom no different from the material conditional, and that the apparent discrepancies can be explained away. But this is very much a minority position.

But note that *both* claims are of the form 'if p, then q' and both have the *same* antecedent p—Oswald not killing Kennedy—and the *same* consequent q—someone else killing Kennedy.

The only difference between the two claims is that (4) is in the *indicative* mood ('... didn't kill ... did.') while (5) is in the *subjunctive* mood ('... hadn't killed ... would have').

Accordingly, claims like (4) are called indicative conditionals and claims like (5) subjunctive conditionals.

(Sometimes subjunctive conditionals are called 'counterfactual' on the grounds that they imply the falsity of their antecedents. But this terminology can be misleading, given that plenty of indicative conditionals also have antecedents that are pretty sure to be false—(4) would be a case in point.)

8.8 Rational and Metaphysical Changes

Let me say a bit more about the difference between indicative and subjunctive conditionals. (I can only scratch the surface here. The analysis of these constructions is hugely controversial, with a literature stretching to thousands and thousands of articles. There are philosophers who spend their whole lives working on conditionals—indeed there are philosophers who work only on indicative conditionals, and others who work only on subjunctive conditionals.)

Indicative conditionals are to do with rational changes of belief. They tell us what we should believe on learning the antecedent p.

Subjunctive conditionals are to do with metaphysical alternatives. They tell us what difference p would have made to the course of history.

To illustrate how indicative conditionals work, suppose that someone whom you trust whispers in your ear that Lee Harvey Oswald definitely didn't kill President Kennedy. What should you now think?

Well, you know full well that Kennedy was assassinated, and your new information doesn't contradict this. So the obvious conclusion is that there was a different assassin. Thus: 'If Oswald didn't kill Kennedy, then someone else did.'

Now take the corresponding subjunctive conditional. The question now is the difference it would have made to history had Oswald not killed Kennedy, not how such information should impact on your beliefs. And to this question the obvious answer (assuming the Warren Commission was right) is that Kennedy would not have been assassinated. Thus: 'If Oswald hadn't killed Kennedy, then no one else would have.'

When we evaluate indicative conditionals, we add p to *all* our current beliefs, make the minimum adjustments needed to accommodate it, and consider whether q still follows.

But when we evaluate subjunctive conditionals, we proceed differently. We first remove from our current beliefs all those whose truth is a causal consequence of not-p—and only then do we add p with minimal adjustments and consider whether q follows. Since we are concerned with the impact p would have on the course of history, we don't want to reason on the basis of facts that would have been causally altered if p had obtained.

That's why we don't hold onto Kennedy's assassination when we make the *subjunctive* assumption 'if Oswald hadn't killed Kennedy . . .'. Removing Oswald's killing Kennedy removes the cause of Kennedy's assassination.

By contrast, we *do* hold onto Kennedy's assassination when we make the *indicative* assumption 'if Oswald didn't kill Kennedy . . .'. Since we are sure that Kennedy actually was killed, we hang onto this information in evaluating the indicative conditional.

FURTHER READING

Colin Howson and Peter Urbach's *Scientific Reasoning: The Bayesian Approach* (Open Court second edition 1993) shows how 'Bayesianism' illuminates many aspects of scientific reasoning.

Paul Horwich's *Probability and Evidence* (Cambridge University Press 1982) covers much of the same ground.

There is a useful Stanford Encyclopedia entry on Bayesian thinking by William Talbott: <http://plato.stanford.edu/entries/epistemology-bayesian>.

A Philosophical Guide to Conditionals (Oxford University Press 2003) by Jonathan Bennett is a masterly introduction to this complex topic.

Mark Sainsbury's *Logical Forms* (Blackwell second edition 2001) contains much useful material about conditionals and their connection with probabilities.

See also <http://plato.stanford.edu/entries/conditionals> by Dorothy Edgington.

EXERCISES

1. If Pr(wind) = 0.6, Pr(rain) = 0.5, and Pr(wind and rain) = 0.4, what is Pr(wind/rain), and what is Pr(rain/wind)?
2. If I draw one card from a well-shuffled pack, what is the conditional probability of:

   (a) a court cart (A, K, Q, J) given a heart
   (b) a court card given not a heart
   (c) a heart given a court card
   (d) not a heart given a court card
   (e) an even number given a non-court card
   (f) an odd number given a non-court card
   (g) an even number given a court card?
3. Suppose you have good reason to hold that Pr(h) = 0.1, Pr(e) = 0.2, and Pr(e/h) is 0.8. Then you learn e. What probability should you now attach to h?

4. You have a 10% degree of belief that a coin is not fair but has a 75% bias in favour of Heads. You toss it twice and see two Heads. What now should be your degree of belief that it is fair?

5. Which of these conditionals are indicative and which subjunctive?

   (a) If you have visited the moon, then you have forgotten being there.
   (b) If you had visited the moon, then you would have forgotten being there.
   (c) If the British Prime Minister in 2012 were a woman, she would be in disguise.
   (d) If the British Prime Minister in 2012 is a woman, she is in disguise.
   (e) If you have eaten arsenic, then you are dead now.
   (f) If you had eaten arsenic, then you would be dead now.
   (g) If the foundations of Buckingham Palace had crumbled to dust, this wouldn't have made it collapse.
   (h) If the foundations of Buckingham Palace have crumbled to dust, this hasn't made it collapse.

6. Which of the conditionals in the last question are true, and which false?

9

Correlations and Causes

9.1 Probabilistic Independence

We say that p is *probabilistically independent* of q when Pr(p/q) = Pr(p).

In such a case, the probability of p on the assumption that q is no different from the probability of p in general. Assuming q doesn't alter the probability of p.

To illustrate, take the propositions that a card drawn from a pack will be an *honour* (10, Jack, Queen, King, or Ace) and that it will be a *heart*. The former is probabilistically independent of the latter. An honour is no more nor less likely on the assumption that the card is a heart than it is anyway.

Let us check the arithmetic. Pr(honour/heart) is Pr(honour *and* heart)—which is 5/52—divided by Pr(heart)—which is 1/4. So Pr(honour/heart) is 5/13, which is just the same as Pr(honour) itself. As I said, getting a heart doesn't make it any more or less likely that you will get an honour.

Note that p is probabilistically independent of q just in case

(1) Pr(p and q) = Pr(p)Pr(q).

(To see why, remember that

$Pr(p/q) = Pr(p \text{ and } q)/Pr(q)$.

So, if

$Pr(p/q) = Pr(p)$ (that is, p is probabilistically independent of q)

then

$Pr(p \text{ and } q) = Pr(p)Pr(q)$,

and vice versa.)

Probabilistic independence thus means that p and q don't occur together any more (or less) often than you would expect given their separate probabilities of occurrence.

We also now see that probabilistic independence is symmetrical. If p is probabilistically independent of q, then q is probabilistically independent of p.

In our example, we have already seen that getting an honour is probabilistically independent of getting a heart. The probability of an honour isn't altered by getting a heart—it's 5/13 either way.

So by the same coin, getting a heart must be independent of getting an honour—and if you think for a second you'll see that the probability of a heart is indeed not altered by getting an honour—it's 1/4 either way.

Just as getting a heart doesn't make it any more or less likely that you will get an honour, so getting an honour doesn't make it any more or less likely that you will get a heart.

We see that when two results are independent, neither gives any information about the other.

9.2 Probabilistic Dependence

When $Pr(p \text{ and } q) > Pr(p)Pr(q)$, then we say p and q are *positively probabilistically dependent.*

This is equivalent to the requirements that

$Pr(p/q) > Pr(p)$

or that

$Pr(q/p) > Pr(q)$.

In such cases q makes p more likely than it would be otherwise, and p makes q more likely than it would be otherwise.

So for example, getting an honour and getting a 9-or-a-10 are positively probabilistically dependent. The probability of having both (by getting a 10) is 1/13, which is greater than the product of the probabilities of getting an honour (5/13) and getting a 9-or-10 (2/13).

When $Pr(p \text{ and } q) < Pr(p)Pr(q)$—equivalently $Pr(p/q) < Pr(p)$ or $Pr(q/p) < Pr(q)$—then we say p and q are *negatively probabilistically dependent.*

Getting an honour and getting an even numbered card (2, 4, 6, 8, or 10) are negatively probabilistically dependent. The probability of getting both these results (you need a 10 again) is 1/13—which is *less* than the product of the probabilities of getting an honour (5/13) and getting an even-numbered card (1/2).

9.3 Correlation

We speak of correlations when we study the objective probabilistic dependencies between distinct properties of individuals. The individuals might be people, places, countries, cars, stars, cows,... pretty much anything whatever. If we were studying people, our properties might be gender, alcohol consumption, and heart disease, say. If we were studying cows, our properties might be diet, breed, weight, and fertility. And so on.

Suppose we represent the properties of interest in some such case as F, G, H,... We can then use Pr(F), Pr(G), Pr(H),...to represent the objective probability that any given individual will have property F, G, H, ... respectively.

If in such a case F and G are positively probabilistically dependent—$Pr(F/G) > Pr(F)$—then we can say that F and G are *correlated*.

A correlation between F and G thus means that F occurs more often in the presence of G than otherwise (and vice versa). For example, we might find that in people heart disease (H) and drinking alcohol (A) are correlated—$Pr(H/A) > Pr(H)$. This tells us that the probability of heart disease among the alcohol drinkers is higher than in the population in general.[1]

9.4 Causation and Correlation

We're often told that correlation doesn't prove causation. And that's true enough—a craving for ice cream is correlated among women with giving birth some months later, but the craving doesn't cause the birth.

In this case, the correlation isn't due to the craving causing the birth, or vice versa, but to the presence of a common cause for both events—namely, pregnancy. The craving is thus a *symptom* of the impending birth, but not its cause.

Still, even if correlation doesn't always mean causation, because of the possibility of common causes, it is arguable that correlation between two properties does mean that *either* one causes another *or* they have a common cause.

To have a correlation without any such causal explanation would be an absurd general coincidence. Once-off coincidences are only to be

[1] Statistic textbooks will normally give a more complicated definition of correlation, to deal with quantitative properties like weight as well as on-off qualitative properties like gender. But we can ignore quantitative properties here, since they do not affect the basic philosophical points.

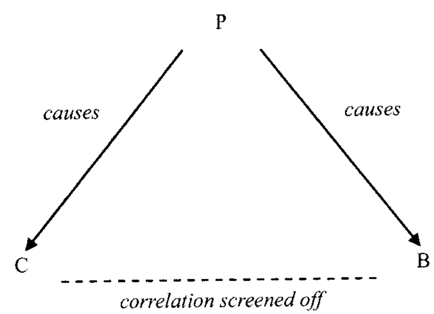
expected. Sometimes Jill and Jane will happen to find themselves wearing the same colour dress just by chance. But if this turns out to be a regular pattern, then it calls for explanation. (Either Jill is copying Jane, or Jane is copying Jill, or they are both influenced by the same fashion advice.)

If we accept that a correlation between two properties does indeed imply that either one is causing the other or that they have a common cause, then we can use this to help us infer causation from correlation. In particular, if we can rule out the possibility of a common cause, then we *can* infer a direct causal connection.

9.5 Screening Off

Interestingly, common causes have a distinctive probabilistic profile. They typically '*screen off*' the correlation between their joint effects, in the sense that this correlation disappears when we '*control for the common cause*'. This allows us to identify common causes from probabilistic patterns, and thereby tell whether or not correlations signify causal connections.

Let me explain this more slowly. Take the craving–birth correlation again. 'Controlling for the common cause' means looking separately



at cases where women are pregnant and where they are not. And, when we do this, the 'correlation disappears' in the sense that, in cases where women are pregnant, the craving for ice cream doesn't now make a subsequent birth any *more* likely, and similarly in cases where women are *not* pregnant. Once we take pregnancy into account, the craving can be seen to make no further difference to the probability of a birth. In this sense, pregnancy '*screens off*' the correlation between cravings and births.

In symbols, there is an initial correlation between craving (C) and birth (B)

$$Pr(B/C) > Pr(B)$$

but this correlation, represented by the dotted line in Diagram 12, is 'screened off' by pregnancy (P) in the sense that:

$$Pr(B/C \textit{ and } P) = Pr(B/P)$$

and

$$Pr(B/C \textit{ and not-}P) = Pr(B/\text{not-}P).$$

Once we know that the craving–birth correlation is 'screened off' by the prior pregnancy in this way, we can infer that there is no causal link between them, and that they are joint effects of pregnancy.

9.6 Spurious Correlations

Of course, we didn't *need* the probabilistic data from the last section to tell us that cravings for ice cream don't cause births. This knowledge is already part of common sense. But in other cases it is precisely such probabilistic data that enable us to find out what is causing what.

To go back to our earlier example, suppose we find that there is a correlation between heart disease (H) and alcohol consumption (A).

$$Pr(H/A) > Pr(H)$$

This might make us think that alcohol consumption causes heart disease. But now suppose that it turns out that gender screens off this correlation—the correlation disappears when we look separately at females (F) and males (not-F).

$$Pr(H/A \textit{ and } F) = Pr(H/F)$$

and

$$Pr(H/A \textit{ and not-}F) = Pr(H/\text{not-}F).$$

This would show that the initial correlation was misleading. Alcohol consumption turns out not to be a genuine cause of heart disease. The two properties are only correlated because gender is a common cause of both. Heart disease tends to be found with alcohol consumption only because being male conduces both to heart disease and to alcohol consumption. (Note that this is just an illustration—I make no claims about its medical accuracy.)

In such a case the original correlation is said to be '*spurious*'. This doesn't mean it is not a real correlation. It is—it is still true that heart disease is more common among the drinkers. But the correlation is spurious in that it doesn't correspond to any direct causal connection—rather the two correlated properties are joint effects of a common cause.

In cases of spurious correlation the common cause is often referred to as a 'confounding' property.

9.7 Randomized Experiments

If we find that some initial correlation between F and G is screened off by some earlier confounding property E, then we can be confident that F and G do not influence each other, but are joint effects of the common cause E, as in the pregnancy and heart disease examples just considered.

However, if we find that some particular earlier E does *not* screen off a correlation between F and G, then we can't immediately infer that G *does* cause F, or vice versa. For there may yet be other common causes we haven't yet identified.

For example, suppose that the heart disease/alcohol consumption correlation turned out *not* to be screened off by gender. We couldn't immediately conclude that alcohol *is* a cause of heart disease. For it may yet be that they are both joint effects of some other 'confounding' property, such as income level, or stress, or anything else—and then drinking would again only be a symptom of this underlying cause, and not itself responsible for heart disease.

The hard way to show that alcohol really is a cause of heart disease is to survey the population and check all the confounding properties that could possibly be responsible for a spurious correlation and show that none of them screens off the association.

But there is an easier way to show that one property is really a cause of another. Suppose we are able to perform a *'randomized experiment'*. The idea here is not to look at correlations in the population at large, but rather to pick out a sample of individuals, and arrange randomly for some to have the putative cause and some not.

The point of such a randomized experiment is to ensure that any correlation between the putative cause and effect *does* indicate a causal connection. This works because the randomization ensures that the putative cause is no longer itself systematically correlated with *any* other properties that exert a causal influence on the putative effect (such as gender, or income level, or stress, ..., or *anything* else). So a remaining correlation between the putative cause and effect must mean that they really are causally connected.

So, for example, we might take a sample of people, and constrain some of them picked at random to drink alcohol and the rest to abstain, in the interests of finding out whether the former group develops more heart disease. Now, of course in this particular case there are

obvious practical and ethical barriers to such an experiment. But in other cases it will be feasible.

Thus suppose we want to make sure that the correlation between some medical treatment and recovery from the relevant disease isn't just a spurious result of the treatment being available only to more affluent sufferers, say, or to some other confounding property. The standard solution is to perform a *'randomized clinical trial'* by taking a group of sufferers and giving the treatment only to a subgroup chosen at random. Many medical experts feel that such randomized trials are the only good way to ascertain the efficacy of medical treatments. (See Box 19.)

9.8 Survey Research

Randomization is a very good way of demonstrating causation. But it is a mistake, notwithstanding the opinion of many in the medical establishment, to suppose that it is the *only* way. Sometimes it is simply not possible, for ethical or practical reasons, to conduct a randomized trial. Then we have to find out about causes the hard way. We need laboriously to survey the overall population and gather data on the correlation between putative cause and effect within subgroups of the population divided by gender, and income level, and stress, ... and all the other things that could possibly be producing a spurious correlation. If none of these screens off the correlation, then this will give us reason to suppose that it reflects a causal connection.

Perhaps we can never be absolutely sure we have checked through every possible confounding factor. But sometimes we can be very confident. We will do well to remember the example of smoking and lung cancer. When the correlation between the two was first noticed, the cigarette companies were quick to suggest that it might be spurious, produced by some common cause like social class, or air pollution, or genetic factors, or...

Now, there was no question of testing this by a randomized trial. (This would have been obviously unethical—you can't take a sample of children and force half of them chosen at random to be smokers.) But this doesn't mean we don't now know that smoking causes cancer. And the way we found out was precisely by surveying all the remotely plausible confounding factors, and showing that none of them in fact screens off the smoking–cancer correlation.

9.9 Simpson's Paradox

Screening off occurs when a common cause is responsible for a positive correlation between two properties even though there is no direct causal connection between them. The lack of a causal connection is exposed by the correlation *disappearing* when we control for the common cause.

There can also be cases where a common cause produces a positive correlation between two properties even though one is in reality a *negative* causal influence on the other. When we control for the common cause the correlation is *reversed*, and what at first looked like a positive cause turns out to have the opposite effect.

Take once more the positive correlation between heart disease (H) and alcohol consumption (A) which initially made it seem that drinking causes heart disease. We earlier supposed that when we controlled for gender and divided the population into females (F) and males (not-F), the correlation would disappear. But now imagine that controlling for gender actually reverses the correlation—that *within* each gender there is *less* heart disease among the drinkers than the rest.

Pr(H/A *and* F) < Pr(H/F)

and

Pr(F/A *and* not-F) < Pr(H/not-F).

Box 19 The Logic of Randomized Trials

In a 'randomized clinical trial' of a medical treatment we take a sample of patients with some ailment and divide them into two groups at random. The 'treatment' group is given the treatment and the 'control' group is not. We then observe whether the recovery rate in the treatment group is significantly higher than in the control group.

The rationale for such trials is to eliminate the danger of spurious correlations. In the wider world, perhaps young people, who are likely to recover anyway, are receiving the treatment more often than old people, and this is creating the impression that the treatment aids recovery. By randomizing the treatment, we forcibly decorrelate it from any such confounding causes as patient age.

Of course, if a treatment does appear efficacious in a particular trial, this could still be due to *statistical fluctuations*. Perhaps by luck the treatment group contained more people who were going to recover anyway. However, this statistical danger is present in any attempt to infer underlying patterns from finite samples, whether or not randomization is involved. And the standard remedy for this statistical danger is to use bigger samples to diminish the probability of misleading fluctuations.

But note that bigger samples are no guard against systematically confounding causes. Suppose that age does indeed influence both recovery and who gets the treatment. Simply getting bigger samples from the population at large isn't going to make this confounding influence go away.

Randomization guards against hidden confounding causes. Big samples guard against statistical fluctuations. Both help to ensure that our inferences are secure.

This would indicate that drinking actually does something to *prevent* heart disease, and only seems initially to cause it because it is more prevalent among men who are prone to heart disease anyway.

This kind of correlation reversal is widely referred to as *'Simpson's paradox'*. But in fact there is nothing terribly paradoxical about such examples. They are quite analogous to ordinary screening off. In both cases, some property appears initially to be a positive cause only because it is itself positively associated with the real cause. The only difference is that in ordinary cases of screening off the putative cause has no real influence at all, whereas in examples of Simpson's 'paradox' it is actually a negative cause.

FURTHER READING

Judea Pearl's *Causality: Models, Reasoning and Inference* (Cambridge University Press 2000) is a detailed study of the relationship between causes and correlations.

There is a useful section on 'Causal Modelling' in Christopher Hitchcock's Stanford Encyclopedia entry on Probabilistic Causation: <http://plato.stanford.edu/entries/causation-probabilistic>.

There is also a Stanford Encyclopedia entry specifically on Simpson's Paradox by Gary Malinas and John Bigelow: <http://plato.stanford.edu/entries/paradox-simpson>.

John Worrall offers an informative critical discussion of the logic of randomized trials in 'Why There's No Cause to Randomize', *The British Journal for the Philosophy of Science* 2007.

EXERCISES

1. When a fair die is thrown, what is the conditional probability of:

   (a) an even number, given a number less than three
   (b) an odd number, given an number greater than three
   (c) a number greater than three, given an odd number
   (d) a number greater than two, given an even number
   (e) a number greater than or equal to two, given a multiple of three
   (f) a multiple of three, given an even number?

2. For each of (a)–(f) in question 1, say whether the two results are independent, positively dependent, or negatively dependent.

3. Which is the odd one out?

   (a) Pr(p/q) > Pr(p)
   (b) Pr(p&q) > Pr(p)Pr(q)
   (c) Pr(not-p/q) > Pr(not-p)
   (d) Pr(not-p & not-q) > Pr(not-p)Pr(not-q)

4. Specify that p and q are probabilistically positively dependent in six different ways.

5. Suppose that the probability of having diabetes (D), being male (M), and being unemployed (U) are given by

   Prob (D) = 0.05
   Prob (M) = 0.60
   Prob (U) = 0.30

   And suppose that

   Prob (D&M) = 0.024
   Prob (U&D) = 0.018
   Prob (U&M) = 0.18

   For each of these last three pairs of properties, say whether the two properties are positively dependent, negatively dependent, or independent. For each of the pairs, work out the conditional probability of the first given the second.

6. Suppose that research shows that Pr(nose cancer/smoking) = 0.3 while Pr(nose cancer) = 0.1.

   Research also shows that:

   Pr(nose cancer/smoking & city-dwelling) = Pr(nose cancer/city-dwelling) = 0.4

   and

   Pr(nose cancer/smoking & country-dwelling) = Pr(nose cancer/country-dwelling) = 0.05.

   What does all this indicate about the causes of nose cancer?

7. Suppose that research in the State University of Euphoria shows that Pr(successful entrance application/male) = 0.4 while Pr(successful entrance application/female) = 0.3.

   Suppose also that the University deals separately with entrance applications to the Arts and Science Faculties, and that further research shows that:

   Pr(successful entrance application/male & Arts) = 0.2 while Pr(successful entrance application/female & Arts) = 0.25

   and

   Pr(successful entrance application/male & Science) = 0.5 while Pr(successful entrance application/female & Science) = 0.6

   What does all this indicate about the factors influencing application success?