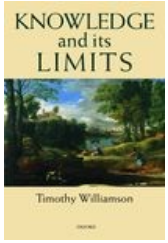

Knowledge and its Limits

Timothy Williamson

https://doi.org/10.1093/019925656X.001.0001

Published: 2002 **Online ISBN:** 9780191598678 **Print ISBN:** 9780199256563

CHAPTER

10 Evidential Probability

Timothy Williamson

https://doi.org/10.1093/019925656X.003.0011 Pages 209–237

Published: October 2002

Abstract

The chapter bases a theory of evidential probability on the equation of knowledge with evidence. It is a form of objective rather than subjective Bayesianism. Updating on new evidence is structured in a way that allows propositions to lose as well as gain probability (forgetting). The account is integrated with possible worlds models of epistemic logic. Since one does not always know what one knows, the accessibility relation is not an equivalence relation, which has the effect that prior probability can diverge from expected posterior probability.

Keywords: Bayesian, epistemic logic, evidential probability, forgetting, possible worlds, updating

Subject: Epistemology, Metaphysics

Collection: Oxford Scholarship Online

10.1 Vague Probability

When we give evidence for our theories, the propositions which we cite as evidence are themselves uncertain. Probabilistic theories of evidence have notorious difficulty in accommodating that obvious fact, as section 9.7 noted. This chapter embeds the fact in a probabilistic theory of evidence. The analysis of uncertainty leads naturally to a simple theory of higher-order probabilities. The first step is to focus on the relevant notion of probability.

Given a scientific hypothesis *h*, we can intelligibly ask: how probable is *h* on present evidence? We are asking how much the evidence tells for or against the hypothesis. We are not asking what objective physical chance or frequency of truth *h* has. A proposed law of nature may be quite improbable on present evidence even though its objective chance of truth is 1. That is quite consistent with the obvious point that the evidence bearing on *h* may include evidence about objective chances or frequencies. Equally, in asking how probable *h* is on present evidence, we are not asking about anyone's actual degree of belief in *h*. Present evidence may tell strongly against *h*, even though everyone is irrationally certain of *h*. We will refer to degrees of belief as

credences; for example, one's prior credence in the proposition that the fair coin will come up heads is normally 1/2; thus credences are *not* the degrees of outright belief discussed in section 4.4.

Is the probability of *h* on our evidence the credence which a perfectly rational being with our evidence would give to *h*? That suggestion comes closer to what is intended, but not close enough. It fails in the way in which counterfactual analyses usually fail, by ignoring side-effects of the conditional's antecedent on the truth-value of the analysandum (Shope 1978). For example, to say that the hypothesis that there are no perfectly rational beings is very probable on our evidence is not to say that a perfectly rational being with our evidence would be very confident that there were no perfectly rational beings. To make the point more carefully, let *p* be a logical truth (a proposition expressed by a logically true sentence) such that in this
p. 210 imperfect world it is very ↳ probable on our evidence that no one has great credence in *p*. There are such logical truths, although in the nature of the case we cannot be confident that we have identified an example. For all we know, they include the proposition that Goldbach's Conjecture is a theorem of first-order Peano Arithmetic (appropriately formalized). Of course, it is not highly probable on our evidence that no one will *ever* give high credence to the proposition that Goldbach's Conjecture is a theorem of first-order Peano arithmetic; we can eternalize the example, if we like, by imagining good evidence that nuclear war is about to end all intelligent life. Let *h* be the hypothesis that no one has great credence in *p*. By assumption, *h* is very probable on our evidence. On the view in question, a perfectly rational being with our evidence would therefore have great credence in *h*. Since *p* is a logical truth, *h* is logically equivalent to the conjunction $p \wedge h$; since a perfectly rational being would have the same credence in logically equivalent hypotheses, it would have great credence in $p \wedge h$. But that is absurd, for $p \wedge h$ is of the Moore-paradoxical form 'A and no one has great credence in the proposition that A'; to have great credence in $p \wedge h$ would therefore be self-defeating and irrational. One can have great credence in a true proposition of that form only by irrationally having greater credence in the conjunction than in its first conjunct. Thus the probability of a hypothesis on our evidence does not always coincide with the credence which a perfectly rational being with our evidence would have in it.

Presumably, a perfectly rational being must give great credence to *p*, be aware of doing so, and therefore give little credence to *h* and so to $p \wedge h$; but then its evidence about its own states would be different from ours. If so, the hypothesis of a perfectly rational being with our evidence is impossible. There is no such thing as *the* credence which a perfectly rational being with our evidence would have in a given proposition. It can be argued that the subjective Bayesian conception of perfect rationality entails perfect accuracy about one's own credences (Milne 1991).

We therefore cannot use decision theory as a guide to evidential probability. Suppose, for example, that anyone whose credences have distribution P is vulnerable to a Dutch Book, a complex bet on which they lose money no matter what the outcome. It may follow that the credences of a perfectly rational being would not have distribution P, if a perfectly rational being would not be vulnerable to a Dutch Book, but it would be fallacious to conclude that probabilities on our evidence do not have distribution P, for those probabilities need not coincide with the hypothetical credences of a perfectly rational being. Perhaps only an imperfectly rational being could have exactly our evidence, which includes our evidence about ourselves. The
p. 211 irrationality of distributing ↳ credences according to the probabilities on one's evidence may simply reflect one's limited rationality, as reflected in one's evidence. But it would be foolish to respond by confining evidential probability to evidence sets which could be the total evidence possessed by a perfectly rational creature. That would largely void the notion of interest; we care about probabilities on *our* evidence.

For all that has been said, any agent with credences which fail to satisfy subjective Bayesian constraints may be *eo ipso* subject to rational criticism. This would apply in particular to the agent's beliefs *about* probabilities on its evidence. But it would apply equally to the agent's beliefs about objective physical chances, or anything else. Just as it implies nothing specific about objective physical chances, so it implies nothing specific about probabilities on evidence.

What, then, *are* probabilities on evidence? We should resist demands for an operational definition; such demands are as damaging in the philosophy of science as they are in science itself. To require mathematicians to give a precise definition of 'set' would be to abolish set theory. Sometimes the best policy is to go ahead and theorize with a vague but powerful notion. One's original intuitive understanding becomes refined as a result, although rarely to the point of a definition in precise pretheoretic terms. That policy will be pursued here. The discussion will assume an initial probability distribution P. P does not represent actual or hypothetical credences. Rather, P measures something like the intrinsic plausibility of hypotheses prior to investigation; this notion of intrinsic plausibility can vary in extension between contexts. P will be assumed to satisfy a standard set of axioms for the probability calculus: $P(p)$ is a non-negative real number for every proposition p; $P(p)= 1$ whenever p is a logical truth; $P(p \vee q)=P(p)+P(q)$ whenever p is inconsistent with q. If $P(q)> 0$, then the conditional probability of p on q, $P(p \mid q)$, is defined as $P(p \wedge q)/P(q)$. $P(p)$ is taken to be defined for all propositions; the standard objection that the subject may never have considered p is irrelevant to the non-subjective probability P. But P is *not* assumed to be syntactically definable. Carnap's programme of inductive logic is moribund. The difference between green and grue is not a formal one.

Consider an analogy. The concept of *possibility* is vague and cannot be defined syntactically. But that does not show that it is spurious. In fact, it is indispensable. Moreover, we know some sharp structural constraints on it: for example, that a disjunction is possible if and only if at least one of its disjuncts is possible. The present suggestion is that probability is in the same boat as possibility, and not too much the worse for that.

p. 212 On the view to be defended here, the probability of a hypothesis h on total evidence e is $P(h \mid e)$. The last chapter gave an account of when a proposition e constitutes one's total evidence. The best that evidence can do for a hypothesis is to entail it (so $P(h \mid e)= 1$); the worst that evidence can do is to be inconsistent with it (so $P(h \mid e)= 0$). Between those extremes, the initial probability distribution provides a continuum of intermediate cases, in which the evidence comes more or less close to requiring or ruling out the hypothesis.

The axioms entail that logically equivalent propositions have the same probability on given evidence. The reason is not that a perfectly rational being would have the same credence in them, for the irrelevance of such beings to evidential probability has already been noted. The axioms are *not* idealizations, false in the real world. Rather, they show what kind of thing we are choosing to study. We are using a notion of probability which (like the notion of incompatibility) is insensitive to differences between logically equivalent propositions. We thereby gain mathematical power and simplicity at the loss of some descriptive detail (for example, in the epistemology of mathematics): a familiar bargain.

The characterization of the prior distribution for evidential probability is blatantly vague. If that seems to disadvantage it with respect to subjective Bayesian credences, which can be more precisely defined in terms of consistent betting behaviour, the contrast in precision disappears in epistemological applications. Given a finite body of evidence e, almost any posterior distribution results from a sufficiently eccentric prior distribution by Bayesian updating on e. Theorems on the 'washing out' of differences between priors by updating on evidence apply only 'in the limit'; they tell us nothing about where we are now (Earman 1992: 137–61 has a sophisticated discussion). Successful Bayesian treatments of specific epistemological problems (for example, Hempel's paradox of the ravens) assume that subjects have 'reasonable' prior distributions. We judge a prior distribution reasonable if it complies with our intuitions about the intrinsic plausibility of hypotheses. This is the same sort of vagueness as infects the present approach, if slightly better hidden.

One strength of Bayesianism is that the mathematical structure of the probability calculus allows it to make illuminating distinctions which other approaches miss and provide a qualitatively fine-grained analysis of epistemological problems, given assumptions about all reasonable prior distributions. That strength is common to subjective and objective Bayesianism, for it depends on the structure of the probability calculus.

p. 213 On the present approach, which can be regarded as a form of objective ↳ Bayesianism, the axioms of probability theory embody substantive claims, as the axioms of set theory do. For example, the restriction of probabilities to real numbers limits the number of gradations in probability to the cardinality of the continuum. Just as the axioms of set theory refine our understanding of sets without reducing to implicit definitions of 'set', so the axioms of probability theory refine our understanding of evidential probability without reducing to implicit definitions of 'evidential probability'.

The remarks above are not intended to smother all doubts about the initial probability distribution. Their aim is to justify the procedure of tentatively postulating such a distribution, in order to see what use can be made of it in developing a theory of evidential probability. That is the focus of this chapter.[1]

10.2 Uncertain Evidence

Suppose that evidential probabilities are indeed probabilities conditional on one's evidence. Then, trivially, the evidence itself has evidential probability 1. P(e| e)= 1 whenever it is defined. Does this require evidence to be absolutely certain? If so, how can evidential probabilities fit into a non-Cartesian epistemology? Section 9.7 gave the problem a preliminary discussion. Let us now consider it more thoroughly.

Section 9.5 defended the assumption that evidence is propositional. Since the approach in this chapter identifies evidential probabilities with probabilities conditional on the evidence, it is in any case committed to treating evidence as propositional. P(h | e)=P($h \wedge e$)/P(e); this equation makes sense only if the evidence e is propositional. We therefore cannot avoid attributing evidential probability 1 to the evidence by denying that evidence is propositional, for then evidential probabilities would be undefined.

We should question the association between evidential probability 1 and absolute certainty. For subjective Bayesians, probability 1 is the highest possible degree of belief, which presumably is absolute certainty. If one's credence in p is 1, one should be willing to accept a bet on which one gains a penny if p is true and is tortured horribly to death if p is false. Few propositions pass that test. Surely complex logical truths do not,
p. 214 even though the probability axioms assign them probability 1. ↳ But since evidential probabilities are not actual or counterfactual credences, why should evidential probability 1 entail absolute certainty?

There is a further link between probability 1 and certainty. Bayesian accounts of learning from experience give a significance to probability 1 which does not depend on any identification of probabilities with actual or counterfactual credences. Suppose that the new evidence gained on some occasion is e. On the standard Bayesian account of this simple case, probabilities should be updated by *conditionalization* on e. The updated unconditional probability of p is its previous probability conditional on e:

$$\text{BCOND } \mathrm{P}_{\text{new}}(p) = \mathrm{P}_{\text{old}}(p|e) = \mathrm{P}_{\text{old}}(p \wedge e)/\mathrm{P}_{\text{old}}(e) \qquad (\mathrm{P}_{\text{old}}(e) \neq 0)$$

We can interpret BCOND as a claim about evidential probabilities. Note that P_{old} is not absolutely prior probability P, but probability on all the evidence gained prior to e. Suppose further, as Bayesians often do, that such conditionalization is the only form of updating which the probabilities undergo. By BCOND, $\mathrm{P}_{new}(e)$= 1. When P_{new} is updated to Pvnew Pnew by conditionalization on still newer evidence f, $\mathrm{P}_{vnew}(e)$ = (e | f)=$\mathrm{P}_{new}(e \wedge f)/\mathrm{P}_{new}(f)$= 1 whenever conditionalization on f is defined. Thus e will retain probability 1 through all further conditionalizations. Since no other form of updating is contemplated, e will retain probability 1. Once a proposition has been evidence, its status is as good as evidence ever after; probability 1 is a lifetime's commitment. On this model of updating, when a proposition becomes evidence it acquires an epistemically privileged feature which it cannot subsequently lose. How can that be? Surely any proposition learnt from experience can in principle be epistemically undermined by further experience.

What propositions could attain that unassailable epistemic status? Science treats as evidence propositions such as ‘Thirteen of the twenty rats injected with the drug died within twenty-four hours’; one may discover tomorrow that a disaffected laboratory technician had substituted dead rats for living ones. The Cartesian move is to find certainty in propositions about one's own current mental state (‘I seem to see a dead rat’; ‘My current degree of belief that thirteen of the twenty rats died is 0.97’). Arguably, we are fallible even about our own current mental states (see Chapters 4 and 8). But even if that point is waived, and we are assumed to be infallible about a mental state when we are in it, we do not remain infallible about it later. However certain I am today of the proposition which I now express by the sentence ‘I seem to see a dead rat’, I may be uncertain tomorrow of the same proposition, then expressed by the sentence ‘Yesterday I
p. 215 seemed to see a dead rat’. I can wonder whether I really remember seeming to see a dead rat, or only ↳ imagine it. Perhaps ‘I seem to see a dead rat’ (uttered by me today) and ‘Yesterday I seemed to see a dead rat’ (uttered by me tomorrow) do not express exactly the same proposition. But if I can think tomorrow the proposition expressed by ‘I seem to see a dead rat’ (uttered by me today), then that proposition can become uncertain for me. If I cannot even think it tomorrow, then the problem is even worse, because I *cannot* retain my evidence. We are uncontroversially fallible about our own past mental states. We are likewise fallible about the mental states of others. You can doubt whether I seem to myself to see a dead rat. Even if I tell you that I seem to myself to see one, you may wonder whether I am lying. Yet science relies on intersubjectively available evidence. Even Bayesian epistemologists assume that evidence is intersubjectively available. Consider, for instance, the arguments that individual differences between prior probability distributions are ‘washed out’ in the long run by conditionalization on accumulating evidence. They typically assume that different individuals are conditionalizing on the *same* evidence. If we start with different prior probabilities, and I conditionalize on evidence about my mental state while you conditionalize on evidence about your mental state, then our posterior probabilities need not converge.

In some cases it can be shown that, although our evidence is different, our beliefs will almost certainly converge on each other because they will almost certainly converge on the truth. For example, if a bag contains ten red or black balls, and we take it in turns to draw a ball with replacement, each observing our own draws and not the other's, and conditionalizing on the results, our posterior probabilities for the number of balls of each colour will almost certainly converge to the same values, even if our prior probabilities are quite different, provided that we both assign non-zero prior probabilities to all eleven possibilities. But even this assumes that our evidence consists of true propositions about the results of the draws, not propositions about our mental states. Where does that assumption come from, on a subjective Bayesian view?

The point generalizes. It is tempting to make a proposition *p* certain for a subject S at a time *t* by attributing a special authority to S's belief at *t* in *p*. But then belief in *p* by other subjects or at other times has a special lack of authority, because it is trumped by S's belief at *t*. For example, to the extent to which eyewitness reports of an event have a special status, non-eyewitness reports are vulnerable to being overturned by them. Thus it is hard to see how *any* empirical proposition could have the intertemporal and intersubjective certainty which the conditionalization account demands of evidence.

p. 216 The standard response is to generalize Bayesian conditionalization to ↳ Jeffrey conditionalization (probability kinematics). For a proposition *p*, in Bayesian conditionalization on *e* ($0 < P_{old}(e) < 1$):

$$(\mathrm{i}) \mathrm{P}_{\mathrm{old}}(p) = \mathrm{P}_{\mathrm{old}}(e)\mathrm{P}_{\mathrm{old}}(p|e) + \mathrm{P}_{\mathrm{old}}(\sim e)\mathrm{P}_{\mathrm{old}}(p|\sim e)$$

$$(\mathrm{ii}) \mathrm{P}_{\mathrm{new}}(p) = \mathrm{P}_{\mathrm{new}}(e)\mathrm{P}_{\mathrm{old}}(p|e) + \mathrm{P}_{\mathrm{new}}(\sim e)\mathrm{P}_{\mathrm{old}}(p|\sim e)$$

For BCOND, the weights $P_{new}(e)$ and $P_{new}(\sim e)$ in (ii) are 1 and 0 respectively. Probabilities conditional on e are unchanged ($P_{new}(p|\, e) = P_{old}(p|\, e)$). What has changed is their weight in determining unconditional probabilities; it has increased from $P_{old}(e)$ to 1. But when experience makes e more probable without making it certain, Jeffrey conditionalization allows us to retain (ii) ((i) is automatic) and make $P_{new}(e)$ larger than $P_{old}(e)$ without making it 1. This increases the weight of probabilities conditional on e at the expense of probabilities conditional on $\sim e$, while giving some weight to both. More generally, experience may cause us to redistribute probability amongst various possibilities, whilst leaving probabilities conditional on those possibilities fixed. Let $\{e_1, \ldots, e_n\}$ be a partition (that is, as a matter of logic, exactly one proposition in the set is true; for mathematical simplicity, infinite partitions are ignored) such that $P_{old}(e_i) > 0$ for each i ($1 \leq i \leq n$). Then P_{new} comes from P_{old} by Jeffrey conditionalization with respect to $\{e_1, \ldots, e_n\}$ if and only if every proposition p satisfies:

$$\text{JCOND } P_{\text{new}}(p) = \Sigma_{1 \leq i \leq n} P_{\text{new}}(e_i) P_{\text{old}}(p|e_i)$$

Bayesian conditionalization is just the special case where $\{e_1, \ldots, e_n\} = \{e, \sim e\}$ and $P_{new}(e) = 1$.

Jeffrey conditionalization cannot reduce probabilities from 1. If $P_{new}(p) = 1$ then $P_{new}(p) = 1$ by JCOND. The idea is rather that no empirical proposition need acquire probability 1 when one learns from experience. On the approach of this chapter, by contrast, evidence must have evidential probability 1, and some empirical propositions must be evidence if evidential probabilities are ever to change. Should the present approach be modified to permit Jeffrey conditionalization?

The updating of evidential probability by Jeffrey conditionalization is hard to integrate with any adequate epistemology, because we have no substantive answer to the question: what should the new weights $P_{new}(e_i)$ be? Indeed, if sufficiently fine partitions are used, any probability distribution P_{new} is the outcome of any probability distribution P_{old} by JCOND, provided only that $P_{new}(p) = 1$ whenever $P_{old}(p) = 1$ and the set of
p. 217 relevant propositions is finite.[2] Arguably, the same applies to ↳ BCOND.[3] But there is a simple schematic answer to the epistemological question 'Which instances of BCOND update evidential probability?': those in which e is one's new evidence. Although that answer immediately raises the further question 'What is one's evidence?', it still constitutes progress, for it divides the theoretical labour, allowing other work in epistemology and in philosophy of science—such as Chapter 9—to provide Bayesianism with its theory of evidence. To the parallel question 'Which instances of JCOND update evidential probability?', no such simple answer will do. Jeffrey conditionalization is not conditionalization on evidence-constituting propositions. Moreover, the weights $P(e_i)$ are highly sensitive to background knowledge. When I see a cloth by candlelight, the new probability that it is green depends on my prior knowledge about its colour, the reliability of my eyesight, and the lighting conditions. Attempts to isolate an evidential input in JCOND have not met with success (see Jeffrey 1975, Field 1978, Garber 1980, and Christensen 1992). Jeffrey conditionalization seems not to admit the kind of articulation which would allow work in other areas of epistemology and of philosophy of science to provide it with a standard of appropriateness for the weights. Without such a standard, an account based on Jeffrey conditionalization promises little epistemological insight.

p. 218 Jeffrey evades the normative question by emphasizing the involuntariness of perceptual beliefs. He denies that sense experience provides *reasons* for belief: it is a mere cause, and none the worse for that (Jeffrey 1983): 184–5). However, normative questions arise even for involuntary beliefs. When the sight of a black cat causes a superstitious man to believe that disaster is about to strike, it may be improbable on his evidence that disaster is about to strike. Although most perceptual beliefs are involuntary, Jeffrey himself is willing to judge them by norms, for he regards Bayesianism as a normative theory, not a descriptive one (Jeffrey 1983: 166–7).

Part of the rationale for Jeffrey conditionalization may also depend on an impoverished theory of propositions. Jeffrey's motivating example involves colour vision in poor light; he argues that no proposition 'expressible in the English language' can 'convey the precise quality of the experience' (1983: 165). Surely no context-independent English sentence conveys the precise quality of the experience. It is much less obvious that in the given context no English sentence with perceptual demonstratives (for example, 'It looks like *that*') can express a proposition which would convey the precise quality of the experience, in the sense that Bayesian conditionalization on it would capture the evidential upshot of the experience (see Christensen 1992, but also section 9.5).

The problem about the certainty of evidence arose from the combination of two claims:

PROPOSITIONALITY The evidential probability of a proposition is its probability conditional on the evidence propositions.

MONOTONICITY Once a proposition has evidential probability 1, it keeps it thereafter.

For PROPOSITIONALITY entails that evidence propositions have evidential probability 1, which by MONOTONICITY implies that they have that status ever after, which is epistemologically implausible. Accounts based on Jeffrey conditionalization retain MONOTONICITY but reject PROPOSITIONALITY; however, they do not yield a nonempty account of evidential probability. A more promising strategy is to retain PROPOSITIONALITY and reject MONOTONICITY. It will be pursued here. PROPOSITIONALITY will henceforth be assumed.

Both BCOND and JCOND allow propositions to acquire probability 1, but not to lose it. They are asymmetric between past and future. Thus a model on which all updating is by Jeffrey or Bayesian conditionalization
p. 219 embodies the empirical assumption that evidence is cumulative, in ↳ the sense of MONOTONICITY. In many cases this assumption is false. Bayesians have forgotten forgetting. I toss a coin, see it land heads, put it back in my pocket and fall asleep; once I wake up I have forgotten how it landed. When I saw it land heads, the proposition *e* that it landed heads was part of my evidence; *e* had probability 1 on my evidence. Once I awake, *e* presumably has probability 1/2 on my evidence. No sequence of Bayesian or Jeffrey conditionalizations produced this change in my evidential probabilities. Yet I have not been irrational. I did my best to memorize the result of the toss, and even tried to write it down, but I could not find a pen, and the drowsiness was overwhelming. Forgetting is not irrational; it is just unfortunate. MONOTONICITY is sometimes a useful idealization; it is not inherent in the nature of rationality.

Information loss has a decision-theoretic interest. Before I fall asleep, I am certain that when I wake up I shall have forgotten how the coin landed (I always forget that kind of thing). I am now happy to accept a bet on which I gain £1 if it lands heads and lose £10 otherwise. Tomorrow I shall be happy to accept a bet on which I lose £5 if it lands heads and gain £6 otherwise. If I make both bets, I lose £4 however it lands. I know now that I am vulnerable to such a diachronic Dutch Book, but what can I do? To avoid it by refusing the first bet is just to turn down a certain £1 (compare Skyrms 1993).[4]

A proposition can lose the status of evidence for me even when in the usual sense I forget nothing. Recall an example from section 9.7. I see one red and one black ball put into an otherwise empty bag, and am asked the probability that on the first ten thousand draws with replacement a red ball is drawn each time. I reply '$1/2^{10,000}$'. Part of my evidence is the proposition *e* that a black ball was put into the bag; my calculation relies on it. Now suppose that on the first ten thousand draws a red ball *is* drawn each time, a contingency which my evidence does not rule out in advance, since its evidential probability is non-zero. But when I have seen it happen, I will rationally come to doubt *e*; I will falsely suspect that the ball only looked black by a trick of the light. Thus *e* will no longer form part of my evidence. The traditionalist claim that the possibility of later doubt shows that *e* never was part of my evidence presupposes an untenably Cartesian epistemology.

p. 220 On standard Bayesian accounts of updating, the only present trace of past evidence is in present probabilities. No separate record is kept of evidence, off which a proposition can be struck. But a theory of evidential probability can keep separate track of evidence and still preserve much of the Bayesian framework.[5] Let P be the prior probability distribution, e_w the conjunction of all old and new evidence for one in a case α, and $P_\alpha(p)$ the evidential probability of a proposition p for one in α. The proposal is that P_α is the conditionalization of P on e_α:

$$\text{ECOND } P_\alpha(p) = P(p|e_\alpha) = P(p \wedge e_\alpha)/P(e_\alpha) \qquad (P(e_\alpha) > 0)$$

ECOND formalizes PROPOSITIONALITY. It allows MONOTONICITY to fail, for if one forgets something between t and a later time t^*, being in cases α and α^* at t and t^* respectively, then e_{α^*} need not entail e_α, so possibly $P_{\alpha^*}(e_\alpha) < 1$ even though $P_\alpha(e_\alpha) = 1$. Thus a proposition can decrease in probability from 1. In that sense, evidence need not be certain.

When no evidence is lost between α and α^*, e_{α^*} is equivalent to $e_\alpha \wedge f$, where f is the conjunction of the new evidence gained in that interval, and ECOND implies that P_{α^*} results from conditionalizing P_α on the new evidence f. Formally, for any proposition p:

$$\begin{aligned} P_{\alpha^*}(p) &= P(p \wedge e_\alpha \wedge f)/P(e_\alpha \wedge f) = (P(p \wedge e_\alpha \wedge f)/P(e_\alpha))/(P(e_\alpha \wedge f)/P(e_\alpha)) \\ &= P_\alpha(p \wedge f)/P_\alpha(f) = P_\alpha(p|f) \end{aligned}$$

BCOND is the special case of ECOND when evidence is cumulative. Thus Bayesian conditionalization can be recovered when needed.

The distribution P is conceptually rather than temporally prior; it need not coincide with P_α for any case α in which some subject is at some time, for P is not a distribution of credences, and the subject may have non-trivial evidence at every time. An incidental advantage of this approach is that it helps with the problem of *old evidence* (Glymour 1980: 85–93, Earman 1992: 119–35, Howson and Urbach 1993: 403–8, and Maher 1996). One would like to say that e confirms h if and only if the conditional probability of h on e is higher
p. 221 than the unconditional ↳ probability of h (compare EV in section 9.2). If e is already part of the evidence then its probability is 1, and the conditional probabilities are identical; yet old evidence does sometimes confirm hypotheses. Appeals are sometimes made to probabilities in past or counterfactual circumstances in which the evidence does not include e, but they produce anomalous results, because the evidence in those circumstances may be distorted by irrelevant factors.

Example: a coin is tossed ten times. Let h be the hypothesis that it landed the same way each time. The initial probability of h is $1/2^9$. Witness A says 'I saw the first six tosses; it landed heads each time'. Witness B then says 'I saw the last four tosses; it landed tails each time'; let e be the proposition that B said this. We have no reason to doubt A and B; if they are both telling the truth, then h is false. But B's statement causes A to break down; he admits that he was lying, and has no relevant knowledge. If B had not made his statement, A would not have withdrawn his, and there would have been no reason to suspect that he was lying. Thus, in the nearest past or counterfactual circumstances in which e was not part of our evidence, the conditional evidential probability of h on e is lower than the unconditional evidential probability of h. Nevertheless, in our present situation, e does confirm h, for since we still have no reason to doubt B, the probability of h on our evidence is around $1/2^6$. Once we have the prior probability distribution P, we can say that $P(h \mid e) > P(h)$. If we like, we can relativize confirmation to background information f by requiring that $P(h \mid e \wedge f) > P(h \mid f)$, but this does not justify subjecting it to the vagaries of the evidence we once or would have had. Of course, these remarks are schematic, but at least the general form of the solution does not

introduce the irrelevant complications consequent on an identification of the probabilities with past or counterfactual credences.

10.3 Evidence and Knowledge

Which propositions are one's evidence? Without a substantive conception of evidence, probabilistic epistemology is empty; in practice, it has taken the existence of such a conception for granted without itself supplying one.

Different conceptions of evidence are compatible with ECOND. Chapter 9 defended the simple, natural proposal that one's evidence is one's body of knowledge. More precisely, one's total evidence e_α in a case α
p. 222 is the conjunction of all the propositions which one knows in $\alpha \downarrow (E = K)$.[6] Here 'one' may refer to an individual or a community. Since evidence can lose probability 1, the defeasibility of knowledge by later evidence is no objection to E = K. When I see the black ball put into the bag, the proposition that a black ball was put into the bag becomes part of my evidence because I know that a black ball was put into the bag. When I have seen a red ball drawn each time on the first ten thousand draws, that further evidence undermines my knowledge that a black ball was put into the bag, and the previously known proposition ceases to be part of my evidence. Since only true propositions are known, evidence consists entirely of true propositions, but one true proposition can cast doubt on another.

Subjective Bayesians might identify one's evidence with one's beliefs (understood as propositions of subjective probability 1) rather than with one's knowledge (E = B). Given E = B, one can manufacture evidence for one's favourite theories by manipulating oneself into a state of certainty about appropriate propositions—for example, that one has just seen one's guru perform a miracle. That does not capture the spirit of the injunction to proportion one's belief to one's evidence.

The positive argument for E = K will not be rehearsed here. The rest of the chapter develops the conjunction of E = K with ECOND as a theory of evidential probabilities, in a way which indicates at least their mutual coherence. The concept *knowledge* is sometimes regarded as a kind of survival from stone-age thinking, to be replaced by probabilistic concepts for the purposes of serious twentieth-century epistemology. That view assumes that the probabilistic concepts do not depend on the concept *knowledge*. If E = K and ECOND are true, that assumption is false. The concepts *knowledge* and *evidential probability* are complementary; neither can replace the other.

Some initially surprising results of the theory stem from the point that we are not always in a position to know whether we know something. By E = K, we are not always in a position to know whether something is part of our evidence. Let us briefly rehearse the context in which this consequence is independently plausible. Whether something is part of our evidence does not depend solely on whether we believe it to be part of our evidence. That *p* is part of our evidence is a non-trivial condition; arguably, no non-trivial condition is such that whenever it obtains one is in a position to know that it obtains (see Chapters 4 and 8). But if we are not always in a position to know whether something is part of our evidence, how can we use
p. 223 evidence? We shall sometimes not $\downarrow$ be in a position to know the probability of a proposition on our evidence. How then can we follow the rule 'Proportion your belief in a proposition to its probability on your evidence'?

As noted in earlier chapters, there is a recurrent temptation to suppose that we can follow a rule only if it is always cognitively transparent to us whether we are complying with it. On this view, if we are sometimes not in a position to know whether we are φ-ing when C, then we cannot follow the rule 'φ when C'; at best we can follow the rule 'Do what appears to you to be φ-ing when it appears to you that C'. For instance, we cannot follow the rule 'Add salt when the water boils' because we are not always in a position to know

whether something is really salt, water, or boiling; at best we can follow the rule 'Do what appears to you to be adding salt when what appears to you to be water appears to you to boil'. Can we even follow the modified rule? That something appears to us to be so is itself a non-trivial condition. But we *can* follow the rule 'Add salt when the water boils', even though we occasionally make mistakes in doing so. It is enough that we *often* know whether the condition obtains. Compliance with a non-trivial rule is never a perfectly transparent condition. We use rules about evidence for our beliefs because they are often less opaque than rules about the truth of our beliefs; perfect transparency is neither possible nor necessary.

Just as we can follow the rule 'Add salt when the water boils', so we can follow the rule 'Proportion your belief in a proposition to its probability on your evidence'. Although we are sometimes reasonably mistaken or uncertain as to what our evidence is and how probable a proposition is on it, we often enough know enough about both to be able to follow the rule. It is easier to follow than 'Believe a proposition if it is true', but not perfectly easy. And just as adding salt when the water boils is not equivalent to doing one's rational utmost to add salt when the water boils, so proportioning one's belief in a proposition to its probability on one's evidence is not equivalent to doing one's rational utmost to proportion one's belief in a proposition to its probability on one's evidence. The content of a rule cannot be reduced to what it is rational to do in attempting to comply with it. Evidential probabilities are not rational credences.

The next task is to develop a formal framework for the combination of E = K with ECOND, by appropriating some ideas from epistemic logic.[7] Within this framework, the failure of cognitive transparency for evidential probabilities will receive a formal analysis.

p. 224

10.4 Epistemic Accessibility

For the sake of familiarity, we may speak of notional *worlds* rather than cases. In order to facilitate discussion of intersubjective knowledge, we do not conceive a world as centred on a subject and a time. Rather, we implicitly specify the epistemic perspective by our choice of an accessibility relation between worlds (see below). We assume a set of mutually exclusive and jointly exhaustive worlds. In a given application, worlds need be specific only in relevant respects. We need not assume that all worlds are metaphysically possible, in the sense that they could really have obtained. A set of all worlds is assumed. The relevant propositions are true or false in each world, and closed under truth-functional combinations. We assume that for each set of worlds, some proposition is true in every world in the set and false in every other world.

Let P be a prior probability distribution as in section 10.1. P is assumed to satisfy the axioms of the probability calculus as stated in terms of worlds. Thus $P(p)= 1$ whenever p is true in every world; $P(p \vee q) = P(p)+P(q)$ whenever p and q are in no world both true. Consequently, if p and q are true in exactly the same worlds, $P(p) = P(q)$.[8] For any set of worlds, some proposition is true at exactly the worlds in the set, and all such propositions are equiprobable; thus the assignment of probabilities to propositions induces a unique assignment of probabilities to set of worlds. Conversely, an assignment of probabilities to sets of worlds induces a unique assignment of probabilities to propositions.

Propositions are known or not known in worlds; propositions about which propositions one knows are true or false in worlds. The account will not assume any general principle about knowledge, except that a proposition is true in any world in which it is known. In particular, it will not assume logical omniscience; if p and q are true in exactly the same worlds, one may know p and not know q. Relative to a subject S and a time t, a world x is *epistemically accessible* ('accessible' for short) from a world w if and only if every proposition which S knows at t in w is true in x. A world is accessible if, for all one knows, one is in it. Since knowledge implies truth, every world is accessible from itself. A proposition p is consistent with propositions $q_1, \ldots, q_n$ if and only if all of p and $q_1, \ldots, q_n$ are true in some world; thus, in a world w, p is

p. 225 *consistent with what one knows* if and only if *p* is true in some world accessible $\hookrightarrow$ from *w* (compare the standard possible worlds semantics for the possibility operator $\Diamond$). Similarly, *p* follows from $q_1, \ldots, q_n$ if and only if *p* is true in every world in which all of $q_1, \ldots, q_n$ are true; thus, in a world *w*, *p follows from what one knows* if and only if *p* is true in every world accessible from *w* (compare the standard possible worlds semantics for the necessity operator $\Box$). Trivially, if one knows a proposition then it follows from what one knows, but the converse may fail, since one need not know that which follows from what one knows.

Now assume ECOND and E = K; in all worlds, evidential probabilities are probabilities conditional on one's evidence and one's evidence is what one knows. Relative to a subject S at a time *t*, for any world *w*, e_w is the conjunction of S's evidence at *t* in *w*. By E = K, e_w is true in all and only the worlds accessible from *w*. P_w is the distribution of evidential probabilities for one in *w*. ECOND says that P_w results from conditionalizing the appropriate prior distribution P on e_w.

When the set of worlds is at most countably infinite, a further natural constraint on P is that it be *regular*, in the sense that $P(p) = 0$ only if *p* is true in no world: the probability distribution does not rule out any world in advance. When there are uncountably many worlds, no probability distribution is regular (infinitesimal probabilities are not being considered here). The most natural prior distributions are those for which there is a finite number *n* of worlds, and $P(p) = m/n$ whenever *p* is true in exactly *m* worlds, but such uniformity in P will not be assumed. Since knowledge entails truth, e_w is always true in *w*. Thus when P is regular, $P(e_w) > 0$ for each *w*, so probabilities conditional on e_w are well defined and ECOND defines evidential probabilities everywhere. Regularity also entails that the evidential probability of *p* is 1 only if *p* follows from one's evidence, for if *p* is false in some world in which e_w is true, then $P(\sim p \wedge e_w) > 0$, so $P_w(p) < 1$. Regularity likewise entails that *p* follows from what one knows if and only if the evidential probability of *p* is 1, and that *p* is consistent with what one knows if and only if the evidential probability of *p* is non-zero.

Propositions about evidential probability are themselves true or false in worlds. For example, the proposition that *p* is more probable than not on the evidence is true in *w* if and only if $P_w(p) > 1/2$. Thus propositions about evidential probability themselves have probabilities.[9]

p. 226 In the manner of possible worlds semantics, conditions on accessibility $\hookrightarrow$ correspond to conditions on knowledge, which in turn have implications for evidential probabilities. For example, accessibility is transitive if and only if for every proposition *p* in every world, if *p* follows from what one knows then that *p* follows from what one knows itself follows from what one knows (compare the S4 axiom $\Box p \supset \Box \Box p$ in modal logic). The latter condition follows from the notorious 'KK' principle that when one knows *p*, one knows that one knows *p*; it is slightly weaker, but not weak enough to be true, even for all rational subjects (see Chapter 5). For a regular probability distribution, transitivity is equivalent to the condition that when *p* has evidential probability 1, the proposition that *p* has evidential probability 1 itself has evidential probability 1.

Accessibility is symmetric if and only if for every proposition *p* in every world, if *p* is true then that *p* is consistent with what one knows follows from what one knows (compare the Brouwersche axiom $p \supset \Box \Diamond p$). For a regular probability distribution, symmetry is equivalent to the condition that when *p* is true, the proposition that *p* has non-zero evidential probability itself has evidential probability 1. There is good reason to doubt that accessibility is symmetric. Let *x* be a world in which one has ordinary perceptual knowledge that the ball taken from the bag is black. In some world *w*, the ball taken from the bag is red, but freak lighting conditions cause it to look black, and everything which one knows is consistent with the hypothesis that one is in *x*. Thus *x* is accessible from *w*, because every proposition which one knows in *w* is true in *x*; but *w* is not accessible from *x*, because the proposition that the ball taken from the bag is black, which one knows in *x*, is false in *w*. Let *p* be the proposition that the ball taken from the bag is red. In *w*, *p* is true, but that *p* is consistent with what one knows does not follow from what one knows, for what one knows is consistent with the hypothesis that one knows $\sim p$ (see section 8.2 and Humberstone 1988 for related

issues). On a regular probability distribution, the evidential probability in *w* of the proposition that *p* has non-zero evidential probability falls short of 1 in this case.

Such examples depend on less than Cartesian standards for knowledge and evidence; Bayesian epistemology must learn to live with such standards. Moreover, failures of symmetry can result from processing constraints, even when false beliefs are not at issue (see also Shin and Williamson 1994). For a crude example, imagine a creature which knows all the propositions recorded in its memory; we may pretend for simplicity that it is somehow physically impossible for false propositions to be recorded there. Unfortunately, there is no limit to the time taken to deliver propositions from memory to the creature's
p. 227 central processing ↳ unit. Now toadstools are in fact poisonous for the creature, but it has no memory of any proposition relevant to this truth. It wonders whether it knows that toadstools are not poisonous. It searches for relevant memories. At any time, it has recovered no relevant memory, but for all it knows that is merely because the delivery procedure is slow, and in a moment the memory that toadstools are not poisonous will be delivered, in which case it will have known all along that they are not poisonous. Everything which it knows in the actual world *w* is true in a world *x* in which it knows that toadstools are not poisonous; thus *x* is accessible from *w*. But *w* is not accessible from *x*, because something which it knows in *x* (that toadstools are not poisonous) is false in *w*. Although in *w* the proposition *p* that toadstools are poisonous is true, that *p* is consistent with what it knows does not itself follow from what it knows.

Epistemic logic and probability theory are happily married because the posterior probabilities in *w* result from conditionalizing on the set of worlds epistemically accessible from *w*. This idea has become familiar in standard applications of epistemic logic to the concept of *common knowledge* in decision theory and game theory (see for example Fudenberg and Tirole 1991: 541–72). As usual, the proposition that *p* is common knowledge is analysed as the infinite conjunction of *p*, the proposition that everyone knows *p*, the proposition that everyone knows that everyone knows *p*, and so on. Thus the analysis of common knowledge requires an account of knowledge. Something like the framework above is used, with a separate accessibility relation R_S for each agent S but a common prior probability distribution; different agents can have different posterior probabilities in the same world because they have different sets of accessible worlds on which to conditionalize. 'S knows *p*' (K_S *p*) is given the semantics of '*p* follows from what one knows' with respect to the accessibility relation R_S; thus knowledge is treated as closed under logical consequence (contrast the present account). Furthermore, in decision theory accessibility is usually required to be an equivalence relation (symmetric and transitive as well as reflexive) for each agent. On this model, the agent partitions the set of worlds into a set of mutually exclusive and jointly exhaustive sets. In *w*, the agent knows just those propositions which are true in every world belonging to the same member of the partition as *w*. Informally, imagine that each world presents a particular appearance to the agent, who knows all about appearances and nothing more; thus one world is epistemically accessible from another if and only if they have exactly the same appearance, which is an equivalence relation. The corresponding propositional
p. 228 logic of knowledge is the modal system S5, with K_S in ↳ place of □; one can axiomatize it by taking as axioms all truth-functional tautologies and formulas of the forms $K_S(A \supset B) \supset (K_SA \supset K_SB)$, $K_SA \supset A$, and $\sim K_SA \supset K_S\sim K_SA$, and as rules of inference modus ponens and epistemization (if A is a theorem, so is K_SA). One of the earliest results to be proved on the basis of assumptions tantamount to these was Aumann's '[no] agreeing to disagree' theorem: when the posterior probabilities of *p* for two agents are common knowledge, they are identical (Aumann 1976; the proof relies heavily on the assumption of common prior probabilities).

Earlier examples expose some of the idealizations implicit in the partitional model of knowledge. In particular, the counterexamples to the symmetry of accessibility, and so to the Brouwersche schema $\sim A \supset K_S\sim K_SA$, are equally counterexamples to the S5 schema $\sim K_SA \supset K_S\sim K_SA$, given the uncontentious principle that knowledge implies truth ($K_SA \supset A$). Some progress has been made in generalizing results such as Aumann's to weaker assumptions about knowledge (Bacharach 1985, Geanakoplos 1989, 1992, 1994, Samet 1990, Shin 1993, Basu 1996). It can be argued that, even when logical omniscience is assumed, the

propositional logic of knowledge is not S5 but the modal system KT (alia s T), which one can axiomatize by dropping the axiom schema $\sim K_S A \supset K_S \sim K_S A$ from the axiomatization above (Williamson 1994b: 270–5). What KT assumes about knowledge, in addition to logical omniscience, is just that knowledge implies truth. When $K_S A$ is read as 'It follows from what one knows that A', rather than as 'One knows that A' (where one is S), logical omniscience becomes unproblematic for K_S, whatever S's logical imperfections.

10.5 A Simple Model

We can gain a more intuitive feel for the present account of higher-order probabilities by working through some of its consequences in a toy example. In doing so we can combine it with the account of margins for error in Chapter 5.

According to a straightforward margin for error principle, S knows *p* in a world *w* only if *p* is true in every world sufficiently close to *w* in the relevant respects (which will depend on the particular case). In the simplest models, that condition is sufficient as well as necessary for knowing *p*: K *p* is true in *w* if and only if *p* is true in all worlds close to *w* (for simplicity, we omit the subscript 'S'). Let us introduce an operator B,
p. 229 where B *p* is to mean that *p* is highly probable on S's evidence. On the ↳ present account of evidence, we can then say that in such a model, B *p* is true in *w* if and only if *p* is true in *most* worlds close to *w*. That is only a first approximation, of course, because some worlds close to *w* may be assigned higher probabilities than others, and 'most' is problematic for infinite sets; but we can build the required probability distribution into our understanding of 'most'. The result is a *probabilistic margin for error principle*. Whereas any operator defined by the original margin for error principle is automatically factive, because every world is close to itself, B is not in general factive, because a set which contains most worlds close to *w* need not contain *w* itself. A false proposition can be highly probable on one's evidence; some evidence is misleading. In place of the factiveness principle $K\,p \supset p$, one can expect only the weaker consistency principle $B\,p \supset \sim B \sim p$; two disjoint sets cannot each contain most worlds close to *w*. Contradictories cannot both be highly probable on one's evidence.

For definiteness, we can imagine the worlds of our toy model as forming a two-dimensional infinite grid. For convenience, each world may be identified with a 'point', a pair of coordinates $< x, y >$, where *x* and *y* are any integers. Again for convenience, we may identify propositions with sets of points; a proposition is true at a point if and only if the latter belongs to the former. Let us count the points close to $< x, y >$, the points accessible from it, as just those within one step of it on the grid: $< x, y >$ itself, $< x+1, y >$, $< x-1, y >$, $< x, y+1 >$ and $< x, y-1 >$. All worlds are treated as equiprobable. Let us count most of these five points as in a set if and only if at least four are. Thus the proposition B *p* is true in a world $< x, y >$ if and only if $\{< u, v >: |x-u|+|y-v| \leq 1\} \cap p$ has at least four members, whereas K *p* is true in $< x, y >$ if and only if $\{< u, v >: |x-u|+|y-v| \leq 1\} \cap p$ has five members. We can read B as 'It is at least 80 per cent probable that', understanding 'probable' evidentially.

As an example, let *p* be the proposition {<0,1>, <1,0>, <1,2>, <2,1>}. The only point close to at least four members of *p* is <1,1>, so B *p* is {<1,1>}. No point is close to at least four members of B *p*, so BB *p* is {}. Thus <1,1> is a point at which *p* is false but at least 80 per cent probable, although it is only 20 per cent probable that *p* is at least 80 per cent probable. This illustrates the simultaneous breakdown of factiveness and the BB principle that if *p* is at least 80 per cent probable then it is at least 80 per cent probable that *p* is at least 80 per cent probable. More generally in the model, B is subject to erosion effects typical of margin for error principles. For example, if *p* is any finite set, then $B^k p$ (*k* iterations of B on *p*) is empty for some natural number *k*.[10]

p. 230 As required, B *p* entails $\sim B \sim p$ in the model. Also as we should expect, the closure principle that if $p_1, \ldots, p_n$ logically entail *q* then B $p_1, \ldots,$ B p_n logically entail B *q* holds when $n \leq 1$ but not otherwise. For example, if

p1 is {<0,1>, <1,0>, <1,2>, <2,1>} and p_2 is {<1,0>, <1,1>, <1,2>, <2,1>}, then both B p_1 and B p_2 are {<1,1>}, but $p_1 \wedge p_2$ is the intersection of p_1 and p_2, {<1,0>, <1,2>, <2,1>}; since this has only three members, B($p_1 \wedge p_2$) is {}. Each of p_1 and p_2 is 80 per cent probable at <1,1>, but their conjunction is only 60 per cent probable. By contrast, K satisfies the corresponding closure principle in this model for multi-premise inference.

Another contrast between strict and probabilistic margins for error in this model is that K satisfies the Brouwersche principle $p \supset K{\sim}K{\sim} p$ because closeness is symmetric, but B does not satisfy the corresponding principle $p \supset B{\sim}B{\sim} p$. For example, if p is {<1,1>}, then $\sim B{\sim} p$ is {}, so $B{\sim}B{\sim} p$ is {}.

The exposition of the present theory of probabilities on evidence is now complete, and some readers may wish to skip the rest of this chapter. However, deviations from the partitional model sketched at the end of section 10.4 generate a phenomenon which seems to threaten the proposed marriage of knowledge and probability. The aim of the next section is to understand that phenomenon.

10.6 A Puzzling Phenomenon

The paradoxical phenomenon can be illustrated thus. There are just three worlds: w_1, w_2 and x. As in Figure 3, x is accessible from each world; each of w_1 and w_2 is accessible only from itself. Thus accessibility is reflexive and transitive, but not symmetric. For simplicity, the subject will be treated as logically omniscient; the paradoxical phenomenon does not depend on the failure of knowledge to be deductively closed. Since the only world accessible from x is x itself, if one is in x then one knows that one is in x. Since the worlds accessible from w_i are x and w_i, if one is in w_i then one knows that one is in either w_i or x, but one does not know which; for all one knows, one *knows* that one is in x. In w_i, although one is not in x, and therefore does not know that one is in x, one does not know that one does not know that one is in x. This is
p. 231 just ↳ the failure of the Brouwersche and S5 axioms for knowledge in a non-symmetric model.

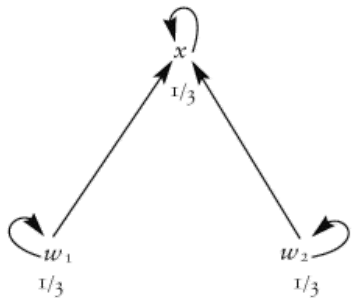

Figure 3

The prior probability distribution is uniform; each world has a prior probability of 1/3. Let p be the proposition that one is in w_1 or w_2. The prior probability of p is 2/3. If one is in x, then p is false in all accessible worlds, so its posterior probability is 0. If one is in w_i, then p is true in just one of the two accessible worlds, so its posterior probability is 1/2. Thus one knows in advance that the posterior probability of p will be either 0 or 1/2, and so in any case lower than its initial probability.[11] But if one knows in advance that, when the evidence comes in, the probability of p on the evidence will drop from 2/3 to at most 1/2, why is that known feature of the future evidence not anticipated by lowering the *prior* probability of p to at most 1/2? Surely the posterior probabilities are a better guide to the truth than the prior probabilities are, because they are based on more evidence (compare Shin 1989 and 1992 and Geanakoplos 1989, 1992, and 1994).

A money pump argument makes the problem vivid. Consider a ticket which entitles one to £6 if *p* is true and to nothing if *p* is false. The initial probability that the ticket entitles one to £6 is 2/3. Given standard Bayesian decision theory, one should be willing to pay up to 2/3£6 + 1/3£0 = £4 in advance for the ticket. But the posterior probability that the ticket entitles one to £6 is at most 1/2, so once the evidence is in one
p. 232 should ↳ be willing to sell the ticket for any price from 1/2£6 + 1/2£0 = £3 upwards. Indeed, if the evidence shows that one is in *x*, then one knows that the ticket is worthless. A shark can apparently pump money out of one by selling one many such tickets for £4 before the evidence is in and buying them back afterwards for £3. I remain a money pump even if I require a small profit on each transaction. Moreover, one knows all that in advance. Is there not something irrational in such an assignment of probabilities?

Reasons emerged in section 10.1 to deny that decision-theoretic arguments have a direct bearing on evidential probabilities. Such arguments are especially dubious when (as above) the probabilities do not all belong to the same time (see, for example, Christensen 1991). Nevertheless, the money pump argument provides an intuitive framework for generalizing the problem. For simplicity, let the worlds form a finite set W. It will be convenient to treat the bearers of probability as subsets of W. For $w \in W$, let R(w) be the set of worlds to which w bears the accessibility relation R. Since e_w is true in exactly the worlds in R(w), the posterior probability $P_w(X)$ of X in w is P(X|R(w)) by ECOND (X ⊆ W). The expectation $E(P_w(X))$ of the evidential probability random variable $P_w(X)$ is therefore $\Sigma\, w{\in}W\, P(\{w\})P(X|R(w))$. The identity of prior and expected posterior probabilities comes to this:

$$\mathrm{EXP}\ \mathrm{P}(\mathrm{X}) = \Sigma_{w \in W} \mathrm{P}(\{w\})\mathrm{P}(\mathrm{X}|\mathrm{R}(w))$$

Consider a ticket which entitles one to £ *n* if one's world is in X and to nothing otherwise. Suppose that before the evidence is in one buys the ticket at its expected (monetary) value at that time; after the evidence is in one sells the ticket at its expected value at that later time. What is one's expected profit or loss over the two transactions? The buying price is P(X)£ *n*. The expected selling price is the expected posterior probability of X times £ *n*. In the example above, the prior probability of *p* was 2/3; its expected posterior probability was 1/3(0) + 2/3(1/2) = 1/3; the expected profit was 1/3 £6 2/3£6, a loss of £2. Thus, if the left-hand side of EXP is less or greater than its right-hand side, one's expected profit over the two transations is positive or negative respectively. Since the prior and expected posterior probabilities of W X are one minus the prior and expected posterior probabilities of X respectively, an expected profit on the two transactions with respect to X implies an expected loss on the corresponding transactions with respect to W X. Thus unless EXP holds, the transactions make one a money pump with respect to some proposition (see Goldstein 1983, Van Fraassen 1984, and Skyrms 1987 for related discussion).

p. 233 One response to the strange situation is to deny that it can arise. On ↳ this view, the money pump argument shows that no probability distribution P on a set of worlds W with an epistemic accessibility relation R can violate EXP for any X ⊆ W; Figure 3 does not picture a genuine possibility. It can be proved that, given a relation R on a finite set W, EXP holds for every regular probability distribution P on W and X ⊆ W if and only if R is an equivalence relation on W (Appendix 4, proposition 5). In partitional models of knowledge, expected posterior probabilities always coincide with prior probabilities; any deviation from partitionality makes them diverge on a suitable probability distribution.[12]

Although the interpretation of R as an accessibility relation for knowledge automatically requires R to be reflexive, one cannot escape the result just by allowing R to be non-reflexive and reinterpreting it as an accessibility relation for (say) rational belief, in the sense that *x* is accessible from *w* if and only if whatever the subject rationally believes in *w* is true in *x*. The aforementioned result holds provided that each member of W has R to at least one member of W, not necessarily itself: in other words, provided that R is serial. The accessibility relation for rational belief is non-serial only when (if ever) rational beliefs are inconsistent. In

that case R(w) is sometimes empty, so the expected posterior probability is not well defined. If R is serial, R(w) is always non-empty; given regularity, the expected posterior probability is then well defined. If the accessibility relation is serial but not reflexive, then expected posterior probabilities diverge from prior probabilities on a suitable probability distribution.

If epistemic accessibility had to be an equivalence relation, EXP would always hold. But the counterexamples to partitionality have not lost their force. Of course, realistic examples involve far more complex epistemic situations than that illustrated above. Nevertheless, we can begin to understand the mechanics underlying non-partitionality by filling out the example above of the worlds w_1, w_2, and x in some detail.

A simple creature monitors the ambient temperature by means of two detectors. When it is not cold, the first detector is activated and causes the information that it is not cold to be stored; otherwise the first detector is inactive. When it is not hot, the second detector is activated and causes the information that it is not hot to
p. 234 be stored; otherwise the second detector is inactive. The relevant three (partial) worlds are w_1 (it is ↳ hot), w_2 (it is cold), and x (it is neither hot nor cold). In w_1, only the information that it is not cold is stored. In w_2, only the information that it is not hot is stored. In x, both the information that it is not hot and the information that it is not cold is stored. Unfortunately, the creature has no capacity to survey what it has stored and detect that a particular piece of information is not stored. Hence, in w_1 it cannot detect that the information that it is not hot is not stored, and infer that it is hot. Similarly, in w_2 it cannot infer that it is cold. Since it never stores false information, we can reasonably treat it as knowing the stored information and no more. Thus the worlds epistemically accessible from w_1 are w_1 and x; the worlds accessible from w_2 are w_2 and x; the only world accessible from x is x itself. Let the three worlds be equiprobable in advance, and treated as such by the creature. Then the epistemic situation is exactly that depicted in Figure 3. If we like, we can elaborate the story to endow the creature with significant powers of logic and self-reflection (see Appendix 5 for details).

Is the initial assignment of equal probabilities to the three worlds irrational? Would some other initial assignment do better? Let P be a regular prior probability distribution which coincides with the corresponding distribution of expected posterior probabilities. So, in particular:

$$\mathrm{P}(\{x\}) = \Sigma_{y \in \mathrm{W}} \mathrm{P}(\{y\}) \mathrm{P}(\{x\} | \mathrm{R}(y))$$

By the diagram $x \in \mathrm{R}(y)$ for all $y \in \mathrm{W}$, so $\mathrm{P}(\{x\}|\mathrm{R}(y)) = \mathrm{P}(\{x\})/\mathrm{P}(\mathrm{R}(y))$. Dividing through by $\mathrm{P}(\{x\})$ gives:

$$1 = \Sigma_{y \in \mathrm{W}} \mathrm{P}(\{y\}) / \mathrm{P}(\mathrm{R}(y))$$

But $\mathrm{R}(x) = \{x\}$, so $\mathrm{P}(\{x\}/\mathrm{P}(\mathrm{R}(x)) = 1$, so:

$$0 = \mathrm{P}(\{w_1\}) / \mathrm{P}(\mathrm{R}(w_1)) + \mathrm{P}(\{w_2\}) / \mathrm{P}(\mathrm{R}(w_2))$$

Thus $\mathrm{P}(\{w_1\}) = \mathrm{P}(\{w_2\}) = 0$. This contradicts the assumed regularity of P. Only an irregular prior distribution on W can coincide with the corresponding expected posterior distribution. Specifically, the proof shows that either $\mathrm{P}(\{x\}) = 0$, in which case $\mathrm{P}(\{x\}|\mathrm{R}(x))$ is undefined, or $\mathrm{P}(\{w_1\}) = \mathrm{P}(\{w_2\}) = 0$. The creature can align its prior probabilities with its expected posterior probabilities only by ruling out some of the three worlds in advance. But that would be quite irrational; each of them is an epistemically live possibility. The uniform prior distribution was not to blame. One must learn to live with the divergence between prior and expected posterior probabilities: but how?

Consider the money pump argument first. As given above, it assumes that, once the evidence is in, the agent
p. 235 can calculate the relevant expectations, ↳ which requires it to know the posterior probabilities. That is just what the structure of the accessibility relation precludes. In the three-world model, it is certain in advance that the posterior probability that it is hot is 1/2 when it is hot and 0 otherwise. Hence if, when it was hot, the creature knew that the posterior probability that it was hot was 1/2, it could deduce that it was hot; but then the posterior probability that it was hot would be 1, not 1/2. For simplicity, sentences about probabilities and actions were omitted from the creature's language; their addition would complicate but not undermine the argument, provided that the creature has no more empirical evidence than before. Thus, when it is hot, the creature cannot know the probability on its evidence that it is hot. It does not know the premises of the decision-theoretic calculation. Even so, the probabilities on its evidence can still play a causal role in its decision-making, for its evidence is physically realized as its stored information. Thus decisions can be made when it is hot that would not have been made if it had not been hot.

Could the creature discover that it is hot by observing its own actions? Once it has acted, it is in a different world; its action may even have changed the temperature. Perhaps it can work out that it *was* hot, but that would not imply that it could have had the present tense knowledge before it acted. Could it have introspected its intention to act in a certain way before carrying it out? Sometimes we do not know whether we are going to act in a certain way until we carry out the action; let the creature be like that when it does not know the probabilities on which it will act.

If we assume that prior probabilities should align themselves with expected probabilities posterior to the future acquisition of knowledge, we assign the probability of being known in the future a privileged status in the present. Why should I give the property of being known by me tomorrow a privileged status today? There is one reason: whatever I shall know tomorrow is true. Thus if I know today that tomorrow I shall know *p*, I can deduce *p* today. By contrast, if I rationally believe today that tomorrow I shall rationally believe *p*, I cannot deduce *p* today; for all I rationally believe today, tomorrow's rational belief will be based on misleading evidence.[13] But this is no reason to give the property of being known by me tomorrow a more privileged status than I give to any other truth-entailing property.

p. 236 Consider an analogy. A die is about to be cast. Each of the natural ↳ numbers from one to six has an equal prior probability (1/6) of being thrown. Exactly five propositions are inscribed on a rock:

e_1 A one will not be thrown.

e_2 A two will not be thrown.

e_3 A three will not be thrown.

e_4 A four will not be thrown.

e_5 A five will not be thrown.

The propositions inscribed on the rock are known to have been chosen at random; that a given proposition is inscribed there does not make it any more likely to be true. Say that a proposition is an *inscribed truth* if and only if it is true and inscribed on the rock; *pseudo-posterior* probabilities are the results of conditionalizing on the conjunction of all inscribed truths. Let *p* be the proposition that a six will be thrown. The prior probability of *p* is 1/6. If a one is thrown, then the inscribed truths are e_2, e_3, e_4, and e_5, so the pseudo-posterior probability of *p* is 1/2. By similar reasoning, the pseudo-posterior probability of *p* is 1/2 if any number between one and five is thrown. If a six is thrown, all the inscribed propositions are inscribed truths, and the pseudo-posterior probability of *p* is 1. Thus the pseudo-posterior probability of *p* is bound to be much higher than its prior probability. Its expected pseudo-posterior probability is (5/6)1/2 +(1/6)1 = 7/12. Pseudo-posterior probabilities are better informed than prior probabilities, because by definition they

result from conditionalizing the latter on true and relevant information. All this is known in advance. Should we therefore revise our prior probabilities to bring them into line with our expected pseudo-posterior probabilities? We have no reason whatsoever to regard a six as any more likely to be thrown than any other number. The inscribed propositions embody a bias towards six. The bias could just as easily have been towards another number, quite independently of the result of the throw.

Moral: it is generally a mistake to try to align one's probabilities with what one knows about the results of conditionalizing them on truths with some given property. One instance of this mistake is to try to align our probabilities with what we know about the results of conditionalizing them on truths which we will know in the future. Although we may be made to suffer for the misalignment, it would not be rational to try to avert the suffering by changing our present beliefs. From our present perspective, the non-partitional structure of
p. 237 our future knowledge is a source of bias, similar in effect to forgetting although much subtler in its ↳ operation. Of course, we shall probably know more tomorrow, and it would be foolish then to disregard the new knowledge. But we cannot take advantage of the new knowledge in advance. We must cross that bridge when we come to it, and accept the consequences of our unfortunate epistemic situation with what composure we can find. Life is hard.

Notes

1 No attempt will be made to survey the non-Bayesian theories of evidential probability in the literature. See e.g. Kyburg 1974 and Plantinga 1993.

2 Proof: Let the propositions of interest be $p_1, \ldots, p_m$. Each of the 2^m possible distributions of truth-values to them corresponds to a conjunction of p_i or $\sim p_i$ for $1 \leq i \leq n$. These 2^m conjunctions form a partition. Perhaps $P_{old}(g) = 0$ for some such conjunction g; by disjoining each such g with a conjunction f such that $P_{old}(f) > 0$, form a partition $\{e_1, \ldots, e_n\}$ such that $P_{old}(e_i) > 0$ for $1 \leq i \leq n$. Each e_i is equivalent to a disjunction $f_i \vee g_i$, where $P_{old}(f_i) > 0$, $P_{old}(g_i) = 0$, and f_i either entails p_j or entails $\sim p_j$ $(1 \leq j \leq n)$. Since $P_{old}(g_i) = 0$, standard reasoning shows that $P_{old}(e_i \equiv f_i) = 1$, so $P_{old}(p_j | e_i) = P_{old}(p_j | f_i)$. Thus $P_{old}(p_j | e_i)$ is 1 or 0, depending on whether f_i entails p_j or $\sim p_j$. Now suppose that for every proposition q, if $P_{old}(q) = 1$ then $P_{new}(q) = 1$. Then $P_{new}(g_i) = 0$, and parallel reasoning shows that if $P^{new}(e_i) > 0$ then $P_{new}(p_j | e_i)$ is 1 or 0, depending on whether f_i entails p_j or $\sim p_j$. Thus if $P^{new}(e_i) \neq 0$, $P_{new}(p_j | e_i) = P_{old}(p_j | e_i)$ $(1 \leq i \leq n)$. From this, JCOND is a routine corollary, with any p_j in place of p.

3 It depends on whether one can introduce finer distinctions than those made by the propositions of interest. If not, and only two possibilities can be distinguished, then no Bayesian conditionalization can change P_{old} to P_{new} where $0 < P_{old}(p) < P_{new}(p) < 1$, because the proposition conditionalized on either makes no difference or eliminates one possibility, in which case all probabilities go to 0 or 1. If finer distinctions can be introduced, $P^{new}(p) = 1$ whenever $P_{old}(p) = 1$, and the set of propositions of interest is finite, then P_{new} comes by BCOND from an extension of P_{old} to the new partition. For suppose that $\{e_1, \ldots, e_n\}$ is a partition such that P_{old} and P_{new} are defined only on propositions equivalent to disjunctions of the e_i. Since $P_{new}(p) = 0$ whenever $P_{old}(p) = 0$, there is a real number $c > 0$ such that for all p for which the probabilities are defined, $cP_{new}(p) \leq P_{old}(p)$. Introduce a new proposition f, bifurcating e_i into $e_i \wedge f$ and $e_i \wedge \sim f$. Determine a probability distribution P_* for the refined partition by: $P_*(e_i \wedge f) = cP_{new}(e_i)$; $P_*(e_i \wedge \sim f) = P_{old}(e_i) - cP_{new}(e_i)$. Hence whenever the probabilities are defined, $P_*(p \wedge f) = cP_{new}(p)$, $P_*(p \wedge \sim f) = P^{old}(p) - cP_{new}(p)$, and $P_*(p) = P_{old}(p)$; thus P_* extends P_{old}. Now $P_*(p | f) = P_*(p \wedge f)/P_*(f) = cP_{new}(p)/c = P_{new}(p)$. Thus P_{new} comes from P_* by BCOND wherever P_{new} is defined. See further Diaconis and Zabell (1982).

4 The case of forgetting shows that not even strong assumptions about the subject's rationality block all counterexamples to the principle of reflection in Van Fraassen 1984. Talbott 1991 has an example of forgetting; the treatment of it by Van Fraassen 1995: 22 cannot plausibly be extended to the present case. For further discussion see Skyrms 1987, Christensen 1991 and 1996, Green and Hitchcock 1994, Howson 1996, Castell 1996, and Hild 1997. Isaac Levi rejects MONOTONICITY in his 1967 and many other publications.

5 Compare the notion of a diary in Skyrms 1983. Skyrms's discussion concentrates on the problem of memory storage, but remembering which propositions are evidence is no worse than remembering a probability for each proposition. Of course, it is often rational to retain a belief even when one has forgotten one's past evidence for it. In some cases the belief itself has attained the status of evidence (see section 10.3); in others one has only indirect evidence for it (for example, one

seems to remember p and is usually right about such things). But even those beliefs are evidentially probable at t only if one's evidence at t supports them. See Harman 1986 for much relevant discussion of clutter avoidance.

6 Restrictive views of evidence can make unnecessary problems for conditionalization by not allowing propositions about the subject's updated belief state to count as part of the new evidence; this may explain the cases discussed in Howson 1996 and Castell 1996.

7 The application of modal logical techniques to epistemological problems was pioneered in Hintikka 1962, although the assumptions made here differ from Hintikka's. A good text for the modal logical background is Hughes and Cresswell 1996.

8 If p and q are true in exactly the same worlds, then in no worlds are both p and $\sim q$ true, and $p \vee \sim q$ is true in all worlds, so by the axioms $1 = P(p \vee \sim q) = P(p) + P(\sim q)$. By the same reasoning, $1 = P(q) + P(\sim q)$. Thus $P(p) = 1 - P(\sim q) = P(q)$.

9 See Skyrms 1980 and Gaifman 1988 for interesting discussions of higher-order probability. Their subjectivism introduces complications into their accounts. For example, Gaifman needs a distinction between the agent's probability and a hypothetical expert's probability to handle higher-order probability. These complications are unnecessary from the present perspective.

10 Proof: $B\,p = \{\langle x, y\rangle : \{\langle u, v\rangle : |x-u| + |y-v| \leq 1\} \cap p \text{ has at least 4 members}\}$. If p is finite then $p\ \{\langle x, y\rangle : |x-a| + |y-b| < k\}$ for some pair $\langle a, b\rangle$ and natural number k. Then if $k-1 \leq |x-a| + |y-b|$, $\langle x, y\rangle \notin B\,p$: for example, if $a \leq x$ and $b \leq y$, then $k \leq |x+1-a| + |y-b|$ and $k \leq |x-a| + |y+1-b|$, so $\langle x+1, y\rangle \notin p$ and $\langle x, y+1\rangle \notin p$, so $\{\langle u, v\rangle : |x-u| + |y-v| \leq 1\} \cap p$ has at most 3 members. Thus $B\,p\ \{\langle x, y\rangle : |x-a| + |y-b| < k-1\}$. By induction, $B^k p = \{\}$.

11 'Know in advance' here could just mean 'know a priori'. As explained in section 10.2, the prior probabilities need not be the evidential probabilities at some earlier time t_0. But the example is more vivid if the initial probabilities are one's evidential probabilities at t_0. On this reading, the diagram does not illustrate one's world at t_0; it is confined to one's worlds at the relevant later time. The problem arises even if the 'prior' probabilities are not the absolutely prior probabilities in ECOND, for the updating can be regarded as an instance of BCOND, which is formally a special case of ECOND.

12 Partitionality is also equivalent to a form of the Principal Principle, or Miller's Principle: $P(X|\{w \in W : P(X|R(w)) = c\}) = c$ for every real number c, $X \subseteq W$, and regular probability distribution P on W such that the conditional probability is defined (Appendix 4, proposition 6). Skyrms 1980 explores the relation of such principles to probability kinematics.

13 The discussion of forgetting in section 10.2 provides one answer to the arguments to the contrary in Van Fraassen 1984 and 1995.