What are conditional probabilities conditional upon?
Keith Hutchison
The British Journal for the Philosophy of Science; Dec 1999; 50, 4; Academic Research Library
pg. 665

What Are Conditional Probabilities Conditional Upon?

Keith Hutchison

ABSTRACT

This paper rejects a traditional epistemic interpretation of conditional probability. Suppose some chance process produces outcomes $X, Y, \ldots$, with probabilities $P(X)$, $P(Y), \ldots$ If later observation reveals that outcome Y has in fact been achieved, then the probability of outcome X *cannot* normally be revised to $P(X|Y)$ $[\equiv P(X \cap Y)/P(Y)]$. This can only be done in exceptional circumstances—when more than just knowledge of Y-ness has been attained.

The primary reason for this failure is that the weight of a piece of evidence varies with the means by which it is provided, so knowledge of Y-ness does not have uniform impact on the probability of X. A better updating of the probability of X is provided by $P(X|Y^*)$, where Y^* is not an outcome of the chance process being observed, but the sentence 'the outcome Y has been observed', an 'outcome' of the subsequent observation process.

This alternative formula is widely endorsed in practice, but not well recognized in theory, where the oversight has generated some unsatisfactory consequences.

1 The epistemic interpretation of conditional probability

The main thesis of this essay is very simple and negative, though supplemented by a more complicated positive sub-thesis. The core of the negative case argues (in Section 3) that there is something dramatically unsatisfactory with the standard epistemic interpretation of conditional probability (the 'Rule of Conditionalization', as it is often called), while the positive coda (Sections 4–5) proposes a replacement, a subtle adaption of the Standard Rule. Though this modified rule is already widely used, the difference between the two formulations is not well recognized. Before the negative case is developed, however, a number of preliminaries need to be dealt with (in Section 2) to clarify the problem—after this long introduction has sketched the road ahead in some detail.

Widely deemed to govern the updating of probabilities (in the various senses of that difficult word) on receipt of new information, the version of the Rule of Conditionalization that I object to is typically formulated as follows:[1]

Suppose circumstances are such as enable the reasonable observer of some chance-process to place 'confidence' $P(X)$ in the outcome being of type X. THEN $P(X|Y)$ $[\equiv P(X \cap Y)/P(Y)]$ represents the confidence that observer should place in the outcome being of type X after the circumstances of the observer change, through addition of just the knowledge that some outcome of type Y has in fact been achieved.

To see what is meant by this Rule, it is useful to choose a concrete example, where there is little confusion about the probability-values, even among those who espouse radically different notions of probability itself. Suppose then a

---

[1] For examples of the epistemic Rule stated in forms that seem equivalent to mine, see Skyrms [1986], p. 189); Resnick ([1986], p. 75); Kyburg ([1990], pp. 50–2); Howson and Urbach ([1993], p. 104); Savage ([1954], p. 44); Maher ([1993], p. 85); Dudewicz and Mishra ([1988], pp. 39–51).

For a version of the rule that differs importantly (yet erratically) from the one I treat as standard here, see my discussion below (in Section **2.3**) of the alternative formulation in terms of H ('hypothesis') and E ('evidence'). The subtle difference in meaning between the outcome 'Y' here, and the vague 'E' there, is a key to this whole essay.

A similar key is provided by the useful treatment in De Finetti ([1974], vol. 1, pp. 134–5). More leisurely than most, this discussion genuinely attempts to explain the rationale behind the Standard Formulation. Two of De Finetti's remarks help to indicate where the view defended in this paper differs from the standard one. Firstly, De Finetti says that 'every evaluation of probability' is conditional upon 'the state of information in which [the observer] finds himself'—but I argue below that it is conditional upon his epistemic *history* (even if he does not know that history). Secondly, De Finetti motivates the Standard Formulation by envisaging a betting game in which every debt is called off 'if the [condition] does not turn out to be true'. This stipulation does not yield *epistemic* conditionals, however. For De Finetti does not tell us what happens when the condition is true but not known to be true. To get the Standard Formulation, one needs to add something omitted by De Finetti, but difficult to comply with: that the debt is to be enforced every time the condition turns out to be true (whether known to be true or not). (An analogous requirement is introduced by in Section **4.2**, when I observe that the Standard Formulation presumes a 'persistent' observation-process.)

For acceptable epistemic conditionals, however, one needs a more accessible rule: the debt has to be enforced if and only if the condition is *known* to be met.

family is chosen from the population of all two-child families in some region, and chosen 'randomly'—in the sense that each family has equal probability of being chosen. (For simplicity, I shall use the word 'random' in this standard sense all through the discussion below.) The probability that this family has two boys in it is approximately $\frac{1}{4}$ (while the probability that it has at least one boy is $\frac{3}{4}$). Suppose that after the choice has been made, it is somehow ascertained that the family has at least one boy in it (without anything else of relevance being discovered). Should the probability that the family has two boys now be updated? If so, has enough information been provided to allow a reliable revision? And if that is so too, how is the updating to be effected?

This example has been chosen to be one of those which precisely fits the Rule of Conditionalization as set out above; and that Rule applies immediately. It answers the first two of our questions in the affirmative; and provides an algorithm that determines the updated probability for us. It gives us indeed no choice but to accept $\frac{1}{3}$ as the updated probability.[2] For this $\frac{1}{3}$ is simply the value here of the formula's $P(X \cap Y)/P(Y)$, with X the outcome ('two boys') of the initial draw whose probability is to be revised, and Y the outcome ascertained ('at least one boy').[3]

Are these three answers good ones, however? I presume we all agree that the first of them is correct.[4] *But the other two answers are very suspect.* Indeed, the dual function of this essay is (firstly) to demonstrate that both are wrong, in that additional information is required to perform the updating reliably; then (secondly) to provide a different, and weaker, rule, that uses the appropriate extra information, to provide a convincing updating. Both these points have been suggested elsewhere (in, say, the works cited in fns 8 and 15)—but with varying degrees of conviction. My task is to purse the question systematically.

To catch a preliminary glimpse of why I take these views, suppose that the existence of a boy is revealed by simply randomly choosing a child from the selected family. Then in a long run of such trials, half the families where a boy is revealed will contain two boys. So it seems that a rational observer, who finds out both *that* and *how* a boy has been detected, should now allocate the value $\frac{1}{2}$ to the probability that the family has two boys. On the other hand, suppose the new knowledge had been acquired quite differently, by someone revealing the existence of a girl whenever possible. Such an informant would only indicate that the family contained a boy if it were impossible to indicate that it contained

---

[2] For some *endorsements* of this answer, see Hodges and Lehmann ([1964], pp. 78–9), plus fn. 7 below. The sources provided in fn. 7, however, need to be read in conjunction with those cited in fn. 9. For some *rejections* of this answer, closely related to those pursued in the present essay (together with pointers to further endorsements of the $\frac{1}{3}$ answer), see fns 6 and 8.

[3] For here: $P(X \cap Y) = \frac{1}{4}$ (the probability-prior-to-the-new-information of the family BOTH containing two boys AND at least one boy), while $P(Y) = \frac{3}{4}$ (the probability-prior-to-the-new-information of the state of affairs actually revealed by that information).

[4] Indeed, my remarks below simply do not address those—propensity-theorists, perhaps?—who believe the new information has no effect on the probability of two boys.

a girl. So after thus finding out that the chosen family contained one boy, the probability that it contained two of them would surely jump to 1.

Whether or not these really are the best answers to our question here has been the subject of much debate (see e.g. the works cited in fns 5, 6, 7 and 8), and it is perhaps not essential to agree with them yet. What immediately matters is: to feel the pull in their direction; to note that they provide a good evaluation of the long-run frequencies; and to accept the clash between them and the posterior probability ($\frac{1}{3}$) calculated using the standard Rule of Conditionalization. Yet that unreliable probability was the one imposed by the traditional algorithm, and we can now see why it is suspect: its observer did not know how the extra sex-information was being supplied. Finding out how the information was provided (apparently) enabled the observer to replace a defective estimate of the updated probability with a better one.[5]

Such conflicting probabilities are very common. Indeed it will turn out that it is only in quite special circumstances that the Standard Formulation of the Rule provides acceptable posterior probabilities. For conditional probabilities of the form $P(X \cap Y)/P(Y)$ in fact serve primarily to tell us something else, *viz.*: 'How are we to update the probability of two boys if we find out that we were wrong about how the family was being chosen, in that it was not chosen at random from the whole population of [two-child] families, but only from a sub-population, that of such families with at least one boy?' (The $\frac{1}{3}$ above is obviously the correct answer to this question.)

The basic problem with the Rule of Conditionalization is that the probability assigned by that Rule (here, the $\frac{1}{3}$) is rarely the same as that to be allocated after *ascertaining* the family belongs to the sub-population, if in truth the random choice was made from the *whole* population. For although arrival of the information does indeed indicate that the family was chosen from the sub-population, it does not provide the slightest reason to believe that it was chosen randomly from that sub-population. (And usually it is not, as I demonstrate at length below, for information is normally—if not always?—biased, and this bias in effect de-randomizes the earlier choice.) So the Rule needs to be reformulated if it is to be successfully applied to the epistemic circumstances set out in its formulation—situations where partial information is received about the outcome of a prior chance-process.

It is certainly not enough to insist that the Standard Rule is only meant to apply to situations where our information arrives free of bias (as hinted by at least one commentator—after glimpsing the inadequacies of the Standard Rule).[6] Firstly because bias is context-dependent, so information seems

---

[5] This is the basic message of Nickerson ([1996]). For other similar messages, see also fns 8 and 15.
[6] *Cf.* Falk ([1993], pp. 56, 177–8), who gives the impression that the probability of two boys in a two-child family after discovering at least one boy would be $\frac{1}{3}$ if the source was not tainted by sexist attitudes.

never to be bias-fee *simpliciter*. Secondly, because it turns out that the Rule only works when there is bias (but the bias has to be appropriate). In consequence, to impose this restriction would be to give up too hastily. For successful conditioning can be routinely effected with biased information.

Indeed, what we need to do instead is just reformulate the Rule given above so that it accords with prevailing practice. For despite the lip-service paid to the Rule in theoretical discussion, epistemic conditionals are often evaluated differently in concrete situations. They are then evaluated using a slightly different rule, inadequately distinguished from the articulation above. This alternative rule happily accommodates biased information; and it uses more data than the Standard Rule, since it typically requires that we know both $P(X\cap Y^*)$ and $P(Y^*)$, where the Y^* is *not* an outcome of the chance-process which takes place prior to the observation. It is instead an outcome of a subsequent observation-process, *viz.*, the epistemic proposition 'the observer has ascertained the original outcome is of type Y'. The revised rule then assigns an updated probability given by $P(X\cap Y^*)/P(Y^*)$.

To see this alternative rule in action, let us return to our family example, supposing that after selecting the family a girl is revealed to the observer whenever possible (i.e. whenever the family contains at least one girl). Otherwise (i.e. when the family contains two boys) a single boy is revealed. The information here is clearly biased against revealing boyness, but that does not hinder us from updating the probability of there being two boys (or girls) in the family, after detecting one boy (or one girl). Indeed, an observer who understands how the information is being released, should allocate a probability of $\frac{1}{3}$ to there being two girls in the family if he finds out there is at least one girl.

This $\frac{1}{3}$ (as we have seen) is the answer given by the Standard Formulation, despite the bias. (It provides, indeed, a preliminary illustration of what I meant above by 'appropriate' bias.) But it is also the answer given by the Y^*-alternative, though I will not pause to perform the calculation, for the Y^*-alternative is more interesting in the case of boy-information. For our observer should obviously reject this standard answer $\frac{1}{3}$ (for the probability of two boys) after finding out *by this means* there is a boy in the family.

Indeed, it is clearly a certainty that there are two boys, if the observer is told there is one. This, however, is the posterior probability provided by the alternative formula. For the Y^*-formula recommends an updating to $\frac{1}{4}/\frac{1}{4}$, since both $P(Y^*)$—the probability of finding out that there is at least one boy—and $P(X\cap Y^*)$—the probability of BOTH there being two boys AND finding out there is a boy in the family—are $\frac{1}{4}$.

Similarly, in our earlier example (where the sex-information was released by

making a random selection), $P(Y^*) = \frac{1}{2}$ and $P(X \cap Y^*) = \frac{1}{4}$, so the alternative formula updates the probability to $\frac{1}{2}$, the figure we earlier recommended (as according with the long-run frequency).

The fact that we have reached updated probabilities here that differ from those provided by the Standard Formulation of the Rule of Conditionalization (and will similarly meet many more below, as summarized in Table 1 in Section **3.3**) does not directly refute that Rule, for more information was used in reaching the non-standard value than the formulation under attack allows. The observer did not just ascertain that there was at least one boy in the family, for he was also presumed to have understood how that information was given to him.

But we have certainly indicated that the updated probability varies with the process by which the information is received, for we have seen that the comparable information had different effects on the updating in the case of girls and boys. So the 'impact' of a piece of information varies with the means by which that information gets acquired. The Standard Rule of Conditionalization gives no hint that this is so, and does worse: it seems to preclude its accommodation. For the Rule's insistence that no additional information reach the observer beyond knowledge that an outcome Y has been achieved means that the pre-conditions for application of the Rule are breached whenever receipt of the new information brings with it the slightest hint as to how it arrived!

We have, however, glimpsed an alternative formula, one which seems capable of accommodating such surplus information. Indeed, the alternative demands it, for its use requires one to know $P(Y^*)$ and $P(X \cap Y^*)$, and these depend on the epistemic procedures. In consequence, the revised rule gives better evaluations of the updated probability.

A thoroughly satisfactory description of the process of conditioning is, however, difficult to formulate. One obstacle to its articulation will tease us all through the discussion below, the apparent need to insist that the Y (or Y^*) here represent the totality of relevant new knowledge—as implied by my including the word 'just' in our initial formulation of the Standard Rule of Conditionalization. Another reflects the fact that it is not, in the end, essential to conditioning that the observer acquire *knowledge* of some *outcome* of the chance-process: one can rationally update probabilities after receiving claims about outcomes that are known to be inaccurate; and one can do the same on receipt of knowledge about matters other than the outcomes of the chance-process under investigation. Such questions will be pursued further below, where they eventually lead us to entertain a revised formula for the posterior probability slightly different to the Y^*-formula above. (We abandon the latter formula only because it is not of wide enough scope, not because it is wrong.)

Another quite different set of difficulties arises because the interpretation of the words and symbols used to express the conditioning process are much at issue—varying with views as to the nature of probability. Some might interpret the 'confidence' as nothing more than an estimate of the long-run frequency; others will see it as a measure of the weight of evidence available to the observer. Some will interpret the X and Y as sentences; others will interpret them as sets, outcomes perhaps of a chance-process. Fortunately, the essence of the case here can be presented without choosing between these alternative interpretations, for the argument develops out of those simple urn models at the centre of our understandings of probability, models where (it seems) the variety of interpretations is of least significance. The case does not involve exotic examples from the fringes of our under-standing—those inconvenient examples where the disputing interpretations drift apart.

Indeed, my focus here will be on core examples of the conditioning process, the cases where there is a minimum of conceptual confusion, where prior probabilities are clearly defined, and where frequency-values (and safe betting ratios) are widely deemed an acceptable guide to prob-abilities. For an attack on the Standard Formulation is all the stronger, the more it relies on unproblematic chance-processes. Accordingly, it is unlikely that the probability values I cite below will be resisted, for (as already hinted) my primary target here is not false estimates of probability. What I seek is a better analysis of the procedures we use to reach those probabilities.

As a consequence of this emphasis on articulation, I am not directly interested here in recommending novel probability assignments, except in the circumstances where too literal a reading of the Rule has triumphed over common sense. A good example of such a retreat from common sense is provided by the widespread claim that the probability of our randomly chosen two-child family having two boys must always change from $\frac{1}{4}$ to $\frac{1}{3}$ upon discovery that the family contains a boy. As already observed, this grossly incorrect result comes from direct application of the Standard Rule of Conditionalization, and (yes!) is sometimes cited as an example of the way theory functions to correct the erroneous common sense of the beginner![7]

---

[7] E.g. Isaac ([1995], pp. 24–5). *Cf.* Feller ([1968], vol.1, pp. 114–8, 125). Note especially the warning (at Feller's p. 116) that these interpretations are not to be taken too seriously, and compare Billingsley ([1986], p. 448). It is not initially clear that these discussions are intended to include epistemic conditionals. To confirm that this is a reasonable reading, one needs also to consult the discussion cited in fn. 9 below.

More fully, my conclusions can be surveyed in the following (overlapping) terms.[8] When $P(Y) \neq 0$,

a) If $P(X|Y)$ is made equal to $P(X \cap Y)/P(Y)$ by fiat (*via* definition, axiom, etc.), then $P(X|Y)$ cannot be given the *epistemic* interpretation need for the Rule of Conditionalization.

b) For $P(X \cap Y)/P(Y)$ is usually not a good estimate of the confidence to be placed in X after discovery that Y is true.

c) Yet $P(X \cap Y)/P(Y)$ is sometimes a good estimate of the confidence to be placed in X after discovery that Y is true—but only in circumstances where more has been ascertained than just the truth of Y. So for $P(X \cap Y)/P(Y)$ to yield the probability posterior to the discovery it is often important that Y *not* be the 'total evidence'.

d) If indeed one knows that the process which provides the knowledge of Y is 'persistent' (in a technical sense elaborated below), then $P(X \cap Y)/P(Y)$ is a good estimate of the confidence to be placed in X after discovery that Y is true.

e) More generally, the confidence to be place in X after ascertaining Y is not a function of the variables X and Y, but is a function of X and a somewhat different variable, one representing the circumstances under which the truth of Y was conveyed to us—Y^* above.

The essence of my case lies in the final claim here, e), for once that is made clear, the other sub-theses will be relatively obvious. In order to establish e), however, we need to confront a conspicuous haziness in the way the traditional Rule of Conditionalization was expressed above, where there was (as already noted) no extended accommodation of the 'principle of total evidence', the idea that Y needs to represent 'all' the evidence that becomes available to us. Though this principle is partially denied in sub-thesis c), a corrected version of it remains essential to all updating of probabilities. yet we also need a version which allows *some* of the new evidence to be ignored—that which is irrelevant. (Otherwise we are quickly led into an infinite regress.) But there are severe difficulties in articulating a criterion for relevance, so the principle

---

[8] The argument of the present paper overlaps somewhat with those in: Bar-Hillel and Falk ([1982]): Freund ([1965]); Korb ([1994]); Nickerson ([1996]); Shafer ([1985]): (plus some further literature cited in these sources). But there are major differences. So Shafer seeks to defend the Standard Formula for $P(X|Y)$, and thus finds a protocol that saves it—the one I refer to below as 'persistent' transmission: Shafer does not seek to emphasize that the formula is false for most protocols; Shafer focuses overwhelmingly on subjective probabilities, and barely appreciates (e.g. p. 269) that the same problem occurs for frequentists. Nickerson and Bar-Hillel and Falk by contrast ignore the Standard Formula; in consequence, they do not develop the contrast between 'ontic' conditionals (to which that formula applies) and epistemic ones; neither do they develop a substitute for that formula. But they fully appreciate that posterior probabilities vary with the means by which a piece of information is received. Korb and Freund certainly recognize this too, but Korb's primary interest lies elsewhere, while Freund presents his argument as an attack on something like the principle of indifference ('the assumption of equal *a posteriori* probabilities', p. 29). See also fn. 15.

remains imperfectly formulated, here and elsewhere. Accordingly, we will continue to proceed *via* concrete example, for once this example is understood, generalization will be reasonably obvious, and a negative case of wide validity can be made without solving the problem of identifying total relevant evidence in general. As already hinted, we eventually see that the relevance of the information is not what matters, in that it is often rational to revise probabilities after the receipt of information which, by normal criteria, is irrelevant.

But before we move to the core of the case, several minor pieces of housekeeping seem to warrant attention.

2 Clearing away some undergrowth

2.1 Non-epistemic interpretations of conditional probability

Conditional probabilities are often interpreted hypothetically, and even counterfactually, $P(X|Y)$ being (say) the probability to allocate X were it to happen to be the case that Y was true. It is important to observe that such 'ontic' interpretations (as we shall call them below) are not being pursued in the present essay—whose concern is with *investigatory* situations, where the truth of Y becomes ascertained or believed, etc. For there are many circumstances in which the two types of conditional probability are dramatically different: Bovine Spongiform Encephalitis can (at the moment of writing) only be diagnosed *via* a post-mortem, so the probability that a cow known to be infected with BSE will be alive tomorrow is zero. But it is widely believed that many living cows have the disease, and can survive for long periods of time with it. The ontic conditional is a dubious guide to the epistemic one here.

Nor are we concerned here with another class of questions, also seen as exemplifying conditional probability, those which estimate the probability of some fact or observation, if a particular scientific theory that bears on the fact is true (and *vice versa*). So I do not ask how we estimate the probability that a cow with BSE will survive one year after infection, if a particular speculation about the operation of BSE happens to be true.

Though such conditionals are not the target of the criticisms below, the discussion does have an indirect bearing on them, in part because of an ongoing tendency in the literature to blur the distinctions between the varieties of conditionals.[9] And more importantly, though the problems we are directly concerned with seem to have a different character, we will eventually conclude that epistemic conditionals are simply special cases of ontic ones—with the

[9] See e.g. Kyburg ([1990] p. 50, where conditional probability is described epistemically, but stochastic independence interpreted ontically); Resnick ([1986], pp. 48, 75, where conditional probability is introduced ontically, then—later—interpreted epistemically, and pp. 53–4, where the author oscillates between ontic and epistemic interpretations); Dudewicz and Mishra ([1988], pp. 38–50, where questions about conditional probability are indiscriminately (?) posed in epistemic and ontic terms).

chance-process specially modified to include the epistemic activity that provides the partial knowledge used for updating. Indeed. it should already be obvious that the *epistemic* conditional which is the probability to be assigned to X when Y has been *ascertained* is just the *ontic* conditional which is the probability to be assigned to X when Y^* is *true*. (This is to be contrasted with $P(X \cap Y)/P(Y)$, the probability to be assigned to X when it is only true that Y is true.)

2.2 The dual chance-processes involved in epistemic conditioning

It is important to note (with e.g. Rosenkrantz [1977], pp. 48–52) that as soon as we pose the sort of question the Rule of Conditionalization is routinely supposed to answer, we are envisaging two quite separate types of activity as taking place. The first is typically a chance-process in which a physical system evolves 'non-deterministically' from some initial macrostate to one of a range of final states—like the throw of a die which ends up in one of six possible orientations on the table below. (The letters X and Y without the asterisks have been used above to refer to final states, 'outcomes', of this initial process.)

The second process is typically subsequent to the first, and is an information-process, conveying a message to some observer that bears on the outcome achieved in the first process. If (say) our die had dots on it of more than one colour, an example of this second type of process would be the discovery that the dots on the uppermost face of the die were all red. As a result of the supplementary information provided by this second process, the observer might well abandon the probabilities he initially allocated to the outcomes of the first process and replace them by updated probabilities. In our case, suppose (to be concrete) that the die had red dots for 1, 2 and 3; and blue dots for 4, 5 and 6. An observer who discovered that red dots had been thrown might then revise the probability of 1 or 2 having been thrown from $\frac{1}{3}$ to $\frac{2}{3}$. Probabilities *prior* to the second process are supplanted by those *posterior* to it. That is what conditioning is about. (More precisely, that is what conditioning *seems* to be about: we eventually see that it is not important that the second process be revealing facts about the outcome of the first one.)

Of course, the second process here, the information-process, may itself involve unpredictable activity, and thus constitute a separate chance-process; and even if the observing activity is not probabilistic, it can still be so conceptualized—since deterministic processes are simply chance ones with probabilities 1 and 0. The things I have designated Y^* above are the outcomes (in the technical Kolmogorov sense) of this second chance-process. So the whole of our discussion here is set paradigmatically in a context where two

distinguishable chance-processes are taking place: an 'evolution-process' (as I shall often call it below), and then an 'observation-process'.

I am not sure that these separate processes are always easy to distinguish, or that the observation has to take place after the evolution. But these conditions are often met, and characterize the conceptually simpler cases of conditioning. Such cases can serve as type-specimens for more complex situations, and accordingly we concentrate on them here. For the fact that the Standard Formulation of the Rule of Conditionalization fails to provide the simplest of probability updatings suffices to rob that formulation of its credibility. To make the negative case, there is no need to look at blurred situations.

It is important to separate the two different processes set out above,[10] because the main problem I detect with the Standard Formulation is its implicit claim that the probabilities associated with the second of the pair, the observation-process, are irrelevant to the updating of probability.[11] Its formulations routinely assert that posterior probabilities can be calculated from nothing more than a knowledge of the probabilities associated with the first process, the temporal evolution, the one that undergoes observation. For they declare that if the second process detects that the first process has produced an outcome of type Y, then the new probability for an outcome of type X is simply $P(X \cap Y)/P(Y)$, where $P(X \cap Y)$ and $P(Y)$ are both probabilities of outcomes of the first process.

In contrast, I claim we need to attend to the probabilities associated with the outcomes Y^* of the second chance-process, the $P(Y^*)$ as I have called them, the probabilities of receiving the conditioning information. The need to embrace these probabilities is widely recognized in practice, for computations of posterior probabilities do in fact often use understandings of the second chance-process (as we will see in the next section). But the fact that it is these probabilities which are being used is often obscured, for the distinction between the two processes is blurred, especially in theoretical discussion.

So the main problem confronted in this discussion is one of articulation, finding the right description of procedures that are already in partial use.

[10] Two of the examples I give in Section **2.3** below of mis-application of the Rule of Conditionalization both involve situations where the first chance-process, that generating the priors, is not specified in any detail, and more or less left to the imagination of the reader. This indeed is one reason that prior probabilities are so often elusive, and (in consequence) so widely distrusted. Yet when the first process is clearly spelled out, the priors are quite unproblematical, and should raise no hackles at all!

[11] For some discussions of conditioning, where it is reasonably clear that the probabilities of the observation-process are overlooked, see: Rosenkrantz ([1977], pp. 48–9, example 1), where the protocol governing the stooge's release of information is not mentioned; Resnick ([1986], p. 56, problem 1), where a 'persistent' process—in the sense set out at in Section **4.2** below—is presumably intended, but no indication is given that this is vital to the discussion; Dudewicz & Mishra ([1988], pp. 38–50), where no hint is given that the mode of observation is relevant. Contrast the discussions cited in fns 8 and 15.

Distinguishing the two chance-process involved in paradigmatic cases of conditioning is a vital first step in this analysis.

2.3 Is the proposed 'new' rule really novel?

There is, however, a popular formulation of the conditioning process that might seem to have solved the problem of articulation here, the formulation phrased in terms of a 'hypothesis' ($H \equiv$ 'the family has two boys') and 'evidence' (E). It assigns an epistemic probability of $P(H \cap E)/P(E)$ to the hypothesis after receipt of the 'evidence', and this formula gives the correct answer—so long as the vague E is interpreted as referring to the second of the dual chance-processes distinguished above. To give the right answer, E has to be short for the sentence 'the evidence is acquired'—so that $P(E)$ is the probability of acquiring the evidence (via that second process).

So in our family example above, where boyness was only revealed if the family had two boys, $P(E)$ can be interpreted as 'the probability of the observation-process indicating there is at least one boy in the family', and this is here $\frac{1}{4}$. So is $P(H \cap E)$. In consequence, the revised probability recommended by this version of the conditioning process is now 1, a satisfactory answer. The calculation simply repeats that already carried out using Y^* above.

In many other situations such an interpretation of E (as referring to the second process) comes naturally, and the formula is repeatedly applied so as to be identical to the rule I formulate using Y^*. It works well—and in so far as it does so, it supports my Y^*-rule. Indeed, it does this so well that it might seem my Y^*-rule adds nothing to it, so the whole discussion here is pointless, targeting a mere straw man.

But this is not so. Firstly, and relatively tritely, because it is not realized that the successful H-E formulation is so different from the standard X-Y formulation set out above. In particular, the successes of the H-E evidence-formulation are not deemed grounds to reject the X-Y outcome-formulation. In consequence, and far more importantly, there are important cases in the literature where this H/E-formulation is used with E referring to Y, the *content* of the evidence, as opposed to Y^*, the *fact* that it is received. So even if it is true that E is *sometimes* interpreted correctly, it is not *always* interpreted correctly, and there is real need to distinguish the Y and Y^* formulations—to avoid a confusion which Korb ([1994], pp. 142–5) aptly calls 'propositional mesmerization'. I give two examples immediately below, both of which have relatively serious consequences, beyond mere quantitative mis-estimation (and Korb, *loc. cit.*, gives another). So the *H/E*-formulation is in fact ambiguous and invites both interpretations, the false and the correct. My aim is to eliminate this source of confusion, by forcing the two interpretations further apart.

A potent example of the failures here is provided by a familiar application of

Bayes' Theorem to situations in which some prediction made by a scientific theory *H* is observed. Howson and Urbach argue ([1993], pp. 119ff) that the probability of the theory is routinely increased by such a confirmation (and many others agree, e.g. Glymour [1980], p. 92 and Salmon [1967], p. 117). But the analyses here are fallacious, because they misidentify the evidence *E*, taking it to be *that* which is observed (my '*Y*'), rather than the *fact* that it is observed (my '*Y**'). For Howson and Urbach take $P(E|H)$ to be 1 (deducing this from the stipulation that $H \Rightarrow E$) and hence quickly infer that the posterior probability of the theory, $P(H|E) = P(H)/P(E)$. This, of course, can never be less than the prior probability $P(H)$, since $P(E) \leq 1$, and will *greatly* exceed $P(H)$ whenever $P(E)$ is small. So they conclude (*op. cit.*, p. 120): 'any evidence [. . .] must confirm every hypothesis [. . .] of which it is a logical consequence' (so long as we stay outside the territory of evidence and theories that are so implausible their probabilities vanish altogether).

To display the fallacy here, let us look at the concrete example provided by our two-child family. We are asking how one should update the probability of the hypothesis *H* (with prior probability $\frac{1}{4}$) that the randomly chosen family has two boys, after observing something that follows from the truth of the hypothesis—that the family has at least one boy.

Let *E* be the sentence expressing the observed consequence of the hypothesis, i.e. 'the family has at least one boy'. Then *E* has prior probability $P(E) = $; and $H \Rightarrow E$. Let further *E** be the fuller statement of the 'evidence', the sentence 'the family is observed to have at least one boy'. A probability cannot be confidently assigned to *E** without more information about the observing process; but it is certainly not true that $H \Rightarrow E^*$.

Because *E** is not entailed by *H*, this *cannot* be what Howson and Urbach (and others who endorse this analysis) mean by the 'evidence'. They must mean something closer to my *E*, and I must assume this is what do mean, as there are no plausible alternatives. Then indeed, the conditional $P(E|H)$ becomes 1 (as they presume).

Given this choice of evidence *E*, however, the posterior probability has to be evaluated as $\frac{1}{3}$, since then $P(H|E) = P(H)/P(E) = \frac{1}{4}/\frac{3}{4}$. Yet this is the answer already noted above as being grossly wrong.

A better treatment would use *E** as the evidence, to yield the posterior probability

$$P(H|E^*) = P(E^*|H).P(H)/P(E^*)$$

which can readily be less than the prior probability. That happens whenever $P(E^*|H) < P(E^*)$. I.e. if some consequence of a hypothesis is especially unlikely to be observed when the hypothesis is true, then witnessing that consequence can decrease the plausibility of the hypothesis—as intuition

would surely suggest.[12] Howson and Urbach and their supporters give an analysis which denies this, for they misrepresent the evidence, misled by the ambiguity of the *H/E*-treatment of conditional probabilities. My target here is no straw man!

We can see much the same phenomenon in a famous discussion of Jeffrey's ([1983], p. 165), centred on the updating of the probability of a hypothesis after arrival of evidence which is so vague that its content cannot be put into words. Jeffrey claims the Rule of Conditionalization does not apply here, because there is no proposition (which I denote '*E*') encapsulating the *content* of the new information acquired. Yet (as Jeffrey himself notes, *op. cit.*, p. 166) there is a proposition (my '*E**') encapsulating the *fact* that the ineffable evidence has arrived—or Jeffrey would not have been able to describe his problem to us. So the applicability of my proposed alternative rule is not called into question. This does not say that the revised formula is correct (or helpful), but it does confirm that it differs from a widely endorsed interpretation of the Standard Formulation of the process of updating. (In fact, the revised rule seems to make Jeffrey's special rule quite redundant, but I do not argue that case here.)

We are finally ready to begin our main critique of that formulation. As already indicated, we use variants of a simple urn experiment, designed to model the two-child family problem—but simplified and elaborated so as to emphasize the role of the observation-processes that provide the updating information.

3 The negative argument

3.1 The basic chance-process

Let us imagine an urn contains (say) eight coins,[13] two with two heads on them, another two with two tails, and the remaining four with one tail and one head. Suppose a totally honest scrutineer ('Eve', we will call her) causes the state of these coins to change in a seemingly 'non-deterministic' manner, by making a *random* draw of one coin from an urn.

---

[12] Before I arrive at work, I am 20% confident that both secretaries will be away (because of a severe epidemic of flu). After I arrive at work I find a note telling me that one of the secretaries is away. But in consequence of finding this note, my confidence that both secretaries are away drops dramatically—because by far the most plausible source for the note is the other secretary.

Let: *H* be the sentence 'both secretaries are away'; *E* be the sentence 'at least one secretary is away'; and *E** be the sentence 'I find a note telling me that at least one secretary is away'. Then $H \Rightarrow E$ (i.e. observation of E = observation of a consequence of H) virtually guarantees that *H* is false.

For another well-known example (due to Richmond Thomason), see Van Fraassen ([1984], p. 246).

[13] At this stage in our analysis the 8 here is not important, and any multiple of 4 (including 4 itself) would serve equally well. The 8 is chosen as the smallest number that will accommodate a variant of the problem introduced below.

Suppose an observer ('Adam') is interested in betting on the head/tail properties of the coin chosen. For such an observer, the draw (our 'evolutionary' chance-process, as articulated in Section **2.2** above) can be well modelled by choosing a 3-member outcome space $\Omega \equiv \{hh, ht, tt\}$, where '*hh*' stands for the outcome-type description 'a two headed coin was chosen', and so on. Since the choice was declared to be random, the function p with values $p(\{hh\}) \equiv p(\{tt\}) \equiv \frac{1}{4}$ & $p(\{ht\}) \equiv \frac{1}{2}$, etc. will suffice to make the pair $<\Omega, p>$ a finite probability space (in the sense of Kolmogorov) that correctly models the chance-process. (I do not defend this claim, simply presuming the reader's familiarity with such matters.)

Let H be the set $\{hh\}$, representing a draw that produces two heads, and let us consider at length the way our observer Adam should update the probability he assigns to this two-head possibility as his epistemic circumstances change. If all Adam knows is that implicit in the specification of the initial draw, he should assign a prior probability to H of $p(\{hh\}) = \frac{1}{4}$.

Suppose Adam later participates in some observation-process, and thus receives additional information about the head/tail-state of the drawn coin, indeed that it has at least one head on it, i.e. that an outcome in the subset $E \equiv \{hh, ht\}$ has been achieved. (Note that $p(E) = \frac{3}{4}$.) If it is also true that this constitutes the totality of his new knowledge, then the traditional formulation of the Rule of Conditionalization is paradigmatically applicable. For we have all the prerequisites for application of that Rule in place: an evolutionary chance-process and an observational chance-process; an acceptable probability model for the evolution-process; and finally, the observation-process has revealed that an event represented by a subset of the outcome space has occurred. *According to the Standard Rule*, this is all we need to know to revise the probabilities the observer should assign to all other outcomes—and it tells us that the new probability to be assigned to H is unequivocally $\frac{1}{3}$.[14]

For reasons that will be spelled out at length soon, I do not accept this answer; and I suspect that many readers (those who reject the analogous answers to the family-with-one-boy problem) already agree with me. For the moment however, that is not what matters. The important point is to accept that one is committed to this answer if one endorses a version of the Rule of Conditionalization like that articulated in the opening of this paper.

A defender of the Standard Rule could, it is true, deny this by insisting it is impossible to find out that the outcome is of type E without finding out something else. Indeed, every time one discovers something about the coin drawn, one will also learn something about oneself, e.g. that one's eyes have recently been open. So E will never represent the total evidence here, and the

---

[14] For that Rule assigns a posterior probability given by the formula $P(H \cap E)/P(E)$ which now takes the values $p(\{hh\} \cap \{hh, ht\})/p(\{hh, ht\}) = p(\{hh\})/\frac{3}{4} = \frac{1}{4}/\frac{3}{4} = \frac{1}{3}$.

conditions for application of the Rule are not met. I do not have any compelling response to this version of the objection, except to remark that it obviously undermines many other versions of the Rule of Conditionalization. If this objection were taken too seriously, it might well protect such rules from refutation, but at the dramatic price of protecting them from application. No sentence could ever be formulated to encapsulate the full evidence, and every case of conditioning would be confronted by Jeffrey's difficulty (noted in Section **2.3**).

One must (I presume) deem that this particular piece of additional knowledge is irrelevant. But as indicated above, I have no useful account to offer of how we decide what is relevant and irrelevant, and hope that it is relatively easy to make this decision in individual cases. In the end, we will observe that relevance is not the big issue, so I am in no hurry to deal with this problem!

3.2 Elaboration of the chance-process

A far more serious version of the same objection observes that detection of an outcome of the evolution-process always brings with it information about some details of the observation-process. One cannot find out that the coin drawn has at least one head on it without understanding *something* about how one found out the coin possessed that head. That 'something' must be accommodated when assessing the total evidence, unless we deem it irrelevant, following the strategy above. But that exit is rightly closed to us here, since an insistence upon the relevance of the observation-process constitutes the germ of my case! So the objection is a valid one.

Yet its impact is slight. For it remains true that the conventional formulation of the Rule of Conditionalization does not force us to accommodate the source of our new knowledge. To see this, let us consider a slight modification of the above chance-process, one designed to leave the evolution-process unaltered in all essentials, but with details of the observation-process spelled out *in advance*. In consequence (as my concrete example should make clear) the observer Adam's understanding of the observation-process is acquired separately from (and well before) the new knowledge about the outcome of the evolution-process—hence cannot now get included in the 'total evidence'.

So let us now suppose that (in this revised process) our scrutineer Eve takes the coin behind a screen after making her draw. There she makes several decisions, *via* processes declared in advance and understood by Adam. After implementing these decisions, she emerges from the screen with the selected coin flat on a saucer but covered. She then shows the saucer to Adam, and removes the cover. His eyes then see the coin, quickly detecting whether it is a head or a tail that is on show.

Prior to Adam's eyes seeing the coin, the probability of there being a

two-headed coin on the saucer is $\frac{1}{4}$. How should this probability be modified just after his eyes identify a head?

The standard version of the Rule of Conditionalization still applies to this revised chance-process, and it commits us to the same answer as before, $\frac{1}{3}$. For an observer interested only in heads and tails can still model the elaborated process by exactly the same probability space as was used to model the simpler process, for the two evolution-processes are effectively identical: they end up with the same three possibilities for the head/tail-state of the coin drawn; and the prior probabilities of these states are quite unchanged.

In this case, however, it is implausible to insist that additional knowledge of relevance was acquired at the time the head was recognized: for Adam possessed a full understanding of how he was going to find out about the coin before his eyes made the identification. The only things he *ascertains* as a result of identifying the coin face is the fact that the coin has at least one head on it—that the outcome is in the subset $\{hh, ht\}$. This knowledge now constitutes the total new evidence of relevance. So the Standard Formulation of the Rule does indeed apply, and it leads to the posterior probability I reject, namely $\frac{1}{3}$.

3.3 Refuting the Standard Formulation

The reason that this posterior probability $\frac{1}{3}$ is unsatisfactory, is that the probability here obviously depends on what decisions the scrutineer Eve implements behind the screen.[15] If she had agreed, for example, to show a tail whenever possible, then she will only show a head when the coin has two heads—and Adam (we have agreed) will know this. So the posterior probability to be allocated by a wise observer to the coin's having two heads when one head has been revealed is 1. Two heads now becomes a certainty after seeing one head.

Not only does this clearly correct answer differ from that provided by the Standard Formula,[16] but more information has been used in its

---

[15] This fact is apparently well recognised by many statisticians, though seems oddly obscured in their discussions. See e.g. Grimmett and Stirzaker ([1993], p. 24, problem 26), but note that the phenomenon is hidden in a problem, while an index search for the critical word 'protocol' leads to no exposition of note. For some other examples of the recognition, see the citations in fn. 8 above.

[16] From oral discussion it has become clear that this claim is controversial. Yet those who dispute it are often unconsciously correcting the Standard Formulation in exactly the direction I recommend. The critical issue is what one takes as the denominator to evaluate the posterior probability, the prior probability of getting at least one head, or the prior probability that one will find out that one has got a head. The formula erroneously takes the former option, for it says that one uses $P(Y)$ the probability of getting (in the evolution-process, the one generating the priors) the discovered outcome Y. This Y is what the observer finds out, so $P(Y)$ is not the probability of discovering Y. An event might have a high probability of taking place, but discovery of the event might be very unlikely.

evaluation than the Standard Formulation deems to be required—more has been used than just the values of $P(X \cap Y)$ and $P(Y)$. Indeed, knowledge restricted to the outcome-probabilities of the evolution-process—the simple selection of the coin—does not here enable Adam to update the probabilities. He cannot reach the answer of the last paragraph using such information alone, for his understanding of the observation-process was clearly vital there. (This is why I earlier described the Standard Formulation as pretending to use too little data, and hence as 'stronger' than the process I articulate.)

The answer provided by the Standard Formulation (*viz.* $\frac{1}{3}$) does, however, apply when an appropriate observation-process is involved. If indeed Eve agrees to show a head whenever possible, the wise observer who understands that this is how she is deciding, should now allocate $\frac{1}{3}$ as the posterior probability. (For a head will then be revealed after $\frac{3}{4}$ of all draws in a long run; but two heads will be present in $\frac{1}{4}$ of draws, all of them included within the where the head is revealed. So the two-heads cases will constitute $\frac{1}{3}$ of the cases where a head is revealed.)

Yet if Eve simply tosses the coin, and reveals the face which comes down uppermost, the situation is more complicated, for the bias on the coins becomes important—even though that (obviously!) had no bearing on the outcome probabilities of the evolution-process, the draw from the urn. *A fortiori*, the bias cannot be reflected in the any conditional probabilities calculated using the Standard Formula. But they can be accommodated in more elaborate calculations that attend to the observation-process.

Indeed, if the coins are known to be unbiased, the posterior probability (for an observer who knows Eve makes her decision by tossing the coin) is easily evaluated as $\frac{1}{2}$. But if the coins are biased (and perhaps differently biased), a whole range of posterior probabilities can be provided, according to the degrees of bias. If the bias is unknown, I do not know what to do—but see Section **4.3** below. Yet if the observer knows the bias, the calculation is straightforward: he uses knowledge of the bias, to calculate how often a head will be seen in a long run; he then calculates in what proportion of those cases, two heads will in fact be on the coin—and takes that proportion to be the posterior probability sought.

So after the outcome Y of some evolution-process has been detected, the posterior probability to be allocated an outcome X depends on more than X, Y, and the prior probabilities of the outcomes of that process. It varies with the epistemic procedure as well. That establishes sub-thesis (e) in my summary above, and constitutes the core of my case. It also explains the title of my essay: for as the table below indicates in summary, conditional probabilities are conditional upon the protocols governing the release of information.

Table 1.

Protocol Used by Honest Scrutineer to Reveal at Least ONE Head on Coin Drawn	Probability to be Accorded to TWO Heads, After Discovering at Least ONE Head
Reveal head only if impossible to reveal tail	100%
Toss (unbiased) coin, then reveal upper face	50%
Reveal head whenever possible	33.3%
Reveal head using complex protocol in Section **4.7**	20%
Reveal head only if coin has head and tail	0%

4 The positive supplement

4.1 Correcting the Standard Formulation

The Standard Formulation of the Rule of Conditionalization claims that after ascertaining that an outcome of type Y has been achieved, one can calculate the posterior probability to be accorded to some other type X, using nothing beyond the range of prior probabilities of all outcomes of the evolution-process. The example just given shows that this is not true, yet makes it fairly obvious how to correct the formula more generally. Clearly, one cannot update the probability without access to the probabilities attached to the observation-processes that provide the updating information about the outcome achieved in the evolution-process under observation. Indeed, the two critical probabilities are: (a) the probability that one will receive the evidence in question, $P(Y^*)$ as I have called it above; and (b) the probability that BOTH some outcome of type X will have been achieved in the process, AND the updating information will have been received by the observer, i.e. $P(X \cap Y^*)$. The posterior probability to be assigned to X is then just the ratio of these two quantities, $P(X \cap Y^*)/P(Y^*)$.[17] This is the replacement formula urged in the opening of the paper.

4.2 When does the Standard Formulation work?

Despite this correction, there remain circumstances in which the Standard Formulation gives an acceptable answer, and it is worth briefly noting some

---

[17] Exactly why this revised formula is true is hard to tell, and the problem of defending it is just a variant of the famous problem of justifying the conventional formulation. See e.g. Howson and Urbach ([1993], pp. 99ff); Maher ([1993], pp. 120ff). A frequency analysis is clear and works well in situations (like my urn draws) where frequencies serve as a reliable guide to probabilities. I do not attempt a more elaborate justification, as I doubt that will be an issue for many readers. They will be unlikely to dispute the rule, merely claiming that it is what is already intended by the Standard Formulation. Hence the detour to confront this question in Section **2.3**.

examples of these. The obvious sufficient (but certainly not necessary: see below) condition for this, is that the observation-process be such as will ensure that both $P(Y) = P(Y^*)$ and $P(X \cap Y) = P(X \cap Y^*)$. These two requirements furthermore will be met (trivially) whenever the observation-process is such as might well be called 'persistent'—meaning that it never fail to inform of the truth of Y if Y is indeed true. So persistence is a sufficient condition for the Standard Formulation—granted, of course, that the observation is also 'reliable' (in the sense of never indicating Y to be true when it is not). Such 'reliability' is implied by the presumption we have been making (right from the beginning of our discussion) that our observer is acquiring *knowledge*, but persistence (in my technical sense) is an additional requirement, and it is routine for epistemic process not to be persistent. Many things are true without our finding out that they are true!

If we think of the observation-process as a *test* (for Y-ness, of course, not X-ness!), then this condition for the accuracy of the Standard Formula is that often expressed by saying that the test produce neither false negatives (my 'persistence') nor false positives ('reliability'). Under such conditions, the Standard Formulation will clearly provide the right answer.

So in the urn draw example, we agreed with the answer $\frac{1}{3}$, if it was true that the scrutineer had agreed to reveal a head whenever possible, i.e. 'persistently'. And in the two-boy family problem, the updated probability $\frac{1}{3}$ seems correct, if the observer knows the family was randomly chosen, and that he has found out that it has at least one boy in it by some persistent procedure, one revealing the existence of a boy whenever a boy is there. Being told that there is a boy in the family only when a girl cannot be revealed, is not such a procedure. That is why the updated probability in these circumstances (*viz.* 1) is not the $\frac{1}{3}$ provided by the formula.

These facts have a further consequence of importance, for they show that the principle of total relevant evidence (as embedded in the Standard Formulation of the Rule of Conditionalization) is false: new knowledge does not have to be restricted to that allowed by this principle; and imposition of such a restriction can generate inaccurate updatings. For we have seen that one can confidently assign the posterior probability $P(X|Y)$ to X, after ascertaining Y, if one has additionally ascertained that the information about Y was provided a persistent and reliable procedure—and if (as well) nothing else of relevance has been ascertained. There is clearly no necessity that the knowledge of the epistemic procedure here be acquired in advance of the knowledge of Y, as the timing of its arrival is not vital. So it is quite reasonable for this knowledge about the *procedure* to arrive slightly after the knowledge about the *outcome* of the evolution process. It would then be part of the new knowledge that creates the need to update probabilities.

New knowledge does not then have to be as restricted as the principle of total

evidence often declares. To assign the posterior probability $P(X|Y)$ to each outcome of type X, after ascertaining that an outcome of type Y has been achieved, it is not necessary for the knowledge that Y has been achieved to represent the *totality* of relevant new knowledge acquired. Highly relevant knowledge—knowledge that the epistemic process was persistent can also be acquired—without jeopardizing the assigned posterior probability.

To relax the rule of total evidence to accommodate this fact is, however, relatively easy, and involves a mere rewrite of our earlier Rule of Conditionalization. Doing this then gives us a restricted version of the Standard Formulation, one that now seems to give acceptable posterior probabilities, albeit within its limited scope. Using italics to indicate the modification, that revision reads:

> Suppose circumstances are such as enable the reasonable observer of some chance-process to place confidence $P(X)$ in the outcome being of type X. THEN $P(X\cap Y)/P(Y)$ represents the confidence that observer should place in the outcome being of type X after the circumstances of the observer change, *through addition of just TWO relevant pieces of knowledge—viz. that some outcome of type* Y *has in fact been achieved, AND that this indication of* Y*-ness was provided by a 'persistent' epistemic process.*

Conversely, if newly arrived knowledge is more restricted—in accordance with the standard version of the principle of total evidence—to knowledge of the outcome-type Y alone, one cannot confidently update the probabilities of the other outcomes. For if one does not already know the procedures, it is not clear how to update the probabilities. It does remains clear, however, that one cannot trust the unadjusted formulation of the Rule of Conditionalization.

But persistence is a rather strong requirement, and incompatible with most everyday epistemic activity (though quite acceptable in the careful model of conditioning that I set out in Section **4.5**). Reliability is (perhaps?) more common, and when information is received reliably, the whole discussion above can be significantly broadened. To illustrate this, I change temporarily to the language of 'hypothesis' and 'evidence', as this seems to make our new situation come alive. For the Standard Rule also gives a correct updating when the truth of some hypothesis H makes no difference to the transmission (as opposed to generation) of a piece of evidence. For $P(H|E)$ then has the same value as $P(H|E^*)$, as can quickly be deduced from our assumption that $P(E^*|E) = P(E^*|H\cap E)$, since reliability ensures that both $P(E^*\cap E) = P(E)$ and $P(E^*\cap H) = P(E^*\cap E \cap H)$. But we still need to know these facts about our epistemic activity to apply that Rule, so again the principle of total evidence has to be relaxed. To update successfully, we need to have grasped our epistemic history.

4.3 Updating in ignorance of the epistemic procedure

To see this last point vividly, we need only consider a variant of the chance-process last set out, where the scrutineer Eve remains honest (so that all information released is 'reliable' in the sense just noted), but where the observer Adam is now denied knowledge of the decision procedures used by Eve, and where, furthermore, she is not obliged to release information after every draw.

Every time his eyes identify a head, the observer Adam will still know with certainty that the coin drawn possesses at least one head, so continues to acquire knowledge about the outcome of the evolution-process, as presumed in the unadjusted Standard Formulation. But he will not now be in a position to assess the weight of his evidence. Eve may, for instance, be refusing to release information when there are two heads on the coin drawn, so that it is impossible for there to be two heads if one head is seen. Conversely, Eve may be doing the exact opposite, showing the head only when there are two heads on the coin: the coin then has to have two heads if the observer sees one head. But Adam cannot distinguish these two diametrically opposed situations. He is in consequence dramatically hampered in allocating posterior probabilities, and can have no confidence in the answer provided by the Standard Formulation.

Adam is not completely incapacitated by these facts, however, and can certainly make guesses based on all sorts of evidence about how he believes Eve is making her decisions, and hence estimate the critical prior probabilities, $\alpha\,[\equiv P(Y^*)$, the probability of his having acquired the knowledge that an outcome of type Y has occurred] and $\beta\,[\equiv P(X \cap Y^*)$, the probability of BOTH his having acquired this knowledge AND an outcome of type X having occurred]. But the posterior probability he will thus assign to two heads (*viz.* β/α) is no more reliable than these guesses, and can be badly discordant with the reality of the situation. I.e. if the scrutineer Eve is choosing by a procedure totally different to that the observer Adam believes she is using, the long-term frequency with which two heads occurs among situations where the single head is revealed can be dramatically different to the probability the observer subjectively holds. Just how entitled he is to hold this probability in such circumstances is a question I do not wish to pronounce upon;[18] but we

---

[18] A subjectivist may insist that since probability represents a degree of belief, the observer has no choice but to hold the 'unsound' probability. But such a subjectivist will have to deny the standard Kolmogorov axiomatization, which attributes a probability of 1 to every tautology. For there are plenty of tautologies so complicated that some observers will not believe they are tautologies.

I do not know how to resolve this dilemma. The positive case I present here seems as strong (and weak) as the claim that a tautology has to be give the probability 1. My negative case is far stronger, for it is clear the Standard Formulation does worse. It will not even allow the observer to hold the conditional probability allocated in the light of his suspicions as to how information is being provided. It insists uniformly on the posterior probability associated with a persistent observation-process.

must certainly recognize that the probability will fail as a guide to safe betting ratios. It is (in some serious sense) a defective estimate of the posterior probability. To update probabilities safely, one needs access to the something beyond that allowed by the unadjusted rule of total evidence—the probabilities governing the release of information.

4.4 Updating upon receipt of irrelevant or incorrect information

The whole of the discussion above has emphasized the importance of distinguishing between Y and Y^*, Y being the knowledge provided to the observer by the action Y^*. We have observed as well that posterior probabilities are directly dependent on the probabilities associated with Y^* rather than Y.

In consequence it can readily happen that it is wise to update probabilities in situations where the Y is irrelevant, so long as the associated Y^* is not irrelevant. *What* the observer finds out may well be of no interest; but the *fact* that he 'hears about it' can be highly significant. Indeed, it can also be wise to update probabilities upon receipt of information Y that is badly wrong (whether suspected to be wrong or not); and it is possible to perform this updating rationally if one knows enough about the process that provided the dubious information. The principles involved in supporting these claims are quite straightforward, and an example should make them clear.

We peruse indeed a further adjustment of the basic urn draw above, attending now to the chemical composition of our eight coins. Suppose that four of the coins happen to be made of bronze (with one bronze coin having two heads; one having two tails; and two having one head and one tail). Suppose also that the remaining four coins are made of nickel. Our observer Adam is to remain interested in betting on the head/tail properties of a randomly drawn coin, and has no interest in the coins' chemical compositions. It is (I take it) obvious that the composition of the coin drawn has no bearing at all on its head/tail-state: the head/tail-state of the coin drawn is 'statistically independent' (to use some common jargon) of its chemical-state. The probability of the selected coin having two heads on it is unaffected by the composition of the coin, and if one finds out that the coin is made of bronze one has found out something that would normally be deemed irrelevant to its head/tail-state.

But this does *not* mean that one should avoid revising the probability of there being two heads on the coin after receiving the information that the coin is bronze.[19] For again, we must look at the observation-process, the rules governing Eve's releasing information. It could happen, for instance, that

---

[19] I.e. two outcomes can be statistically independent in the ontic sense, yet epistemically dependent. The persistent blurring of this distinction in the literature (see fn. 9 above) is simply a variant version of the problem attacked in the present paper *via* the Rule of Conditionalization.

the scrutineer has promised only to releases the information that the coin is bronze when it BOTH has two heads on it AND is bronze. Otherwise, she will release no information about the chemical composition of the coin. Instead she will then indicate that the coin has a tail on it (if the coin does not have two heads), or that it has a head on it (if it has two heads but is not bronze). Under such conditions, the rational observer who knows in advance that this is how information is going to be released will assign a posterior probability of 1 to the coin's having two heads after discovery that it is bronze. Obviously, such a posterior probability is again provided by the formula I recommend, $P(X \cap Y^*)/P(Y^*)$ (for $X \cap Y^* = Y^*$ here). Yet the traditional formulation leads to no change in the probability after discovery that the coin is bronze.[20]

It should also be obvious now that Eve's honesty is not an issue either. It is her practices that matter, not the words she uses to describe those practices, and discord between the words and the actions is of no interest to any observer who understands the actions.[21] Suppose indeed that Eve claims to release information as outlined in the last paragraph, but in fact does something different, revealing a bronze composition only when there is a head and a tail. The observer who knows this should simply ignore what Eve claims to be doing. So if bronze is revealed to him, he will assign a posterior probability of 0 to the coin's possessing two heads. And again, such a posterior probability is recommended by the formula I propose, for now $P(X \cap Y^*)/P(Y^*) = 0$ since in these circumstances $P(X \cap Y^*) = 0$.

Suppose, indeed, that the scrutineer were to make her 'dishonesty' patent, by declaring a familiar falsehood (e.g. that 2 plus 2 is 5) every time there were two heads, but revealed a chemical composition in all other circumstances. The observer who understands her actions is still able to update the probability of two heads, despite the gross irrelevance of her false claims. What matters to him is not the overt information declared in the scrutineer's message, but what might be called the 'latent' information, what one can infer from the fact that the claim has been made.

4.5 A revised model and formula

Given then that the overt content of the message received by the observer is irrelevant to the transformation of prior probabilities into posterior ones, it seems a good idea to reconstrue our paradigm cases of conditioning, by

---

[20] The probability of getting two heads on a bronze coin is 1/8; the probability of getting bronze is 1/2. The conditional probability of getting two heads given a bronze coin is just the ratio of these two probabilities, i.e. 1/4. This is the same as the prior probability of getting two heads.

[21] An observer who does not understand the actions may of course use the words as a guide to the scrutineer's actions, and then honesty, etc. does become important.

imagining that the messages received are particularly bland. This reduces the risk of our being side-tracked by red herrings. So let us suppose that the information provided to the observer (in the observation process) takes an extremely simple form, say the illumination of a lamp, that comes on in one of two colours, green or red. This is intended to provide us with a relatively abstract model of a typical 'test'—to which there are three standard reponses: 'still waiting to hear' (= no light); 'the test was positive' (= green light); 'the test was negative' (= red light).

More fully, I suggest the following type-specimen of the problem of conditioning. A chance-process takes place. A rational observer of the process does not witness its actual outcome, but does attribute probability $P(X)$ to each possible outcome X. After the process has completed, the observer sees a light come on and display the colour green. He receives no other signals of significance. Should he revise the probability he assigns to each outcome X after seeing the green, and if so, how?

Before answering this question, we should briefly note that no generality is lost here if we presume the observer is behaving in such a way as to recognize the green light every time it comes on, i.e. that the information process is persistent in the sense described above.[22] Realizing this enables us to simplify greatly the discussion below, since it means that there is no need to distinguish between epistemic and ontic conditionals. For the sentence 'the green light is on' is true in exactly the same circumstances as the sentence 'the observer realizes that a green light is being displayed'.

It is, of course, obvious that there are circumstances in which the observer should update probabilities after seeing the green signal. It is much less obvious whether there is any general algorithm that governs this updating. A typical situation in which it is desirable to update is when there is a causal link between the unwitnessed outcome of the evolution-process and the colour of the light displayed. The critical factor is the degree of correlation (however

---

[22] For let us suppose this were not so, that the observer was watching a light that shows either blue or yellow, and that he only saw the yellow intermittently. To conditionalize on the non-persistent yellow signal is, however, the same as to conditionalize on an imaginary but persistent green signal that comes on whenever the observer recognizes yellow. So once we invoke the abstraction here of using a simple coloured-light-signal as a standardized representation of the conditionalizing information, there is no need to allow for non-persistent signals. Each of them can be replaced by a persistent one.

In the language of tests, we are presuming that the observer always finds out the result of the test, on the grounds that a test whose results only reach the observer intermittently can be modelled by a different test, one whose results always reach the observer. The modelling test is that variant test which is positive if BOTH the modelled test is positive AND the observer finds this out—negative otherwise.

We are not presuming here that the modelled test itself never produces false negatives. That would require us to make presumptions about more than just how often the green light is seen. We would need to presume that the green light came on every time the property it was *intended* to reveal was present. We cannot judge whether a test produces false negatives (or positives) without knowing the purpose of the test. It is not a property of the test *simpliciter*.

generated) between each outcome X of the evolution-process and the display of
a green light, but to say that is to do little more than rephrase the question
asked.

But at least that rephrasing enables us to focus on what matters, and it is easy
to analyse these in seemingly unproblematic frequentist terms. How often does
the green light come on? In how many of those cases is X true? The updated
probability we seek is surely—but see fn. 17—just the ratio of these two
quantities, i.e. $P(X$ is true & light is green$)/P($light is green$)$.

4.6 Abbreviating the revised model

This 'answer' can be expressed rather differently. We can stop thinking of
conditioning as involving a *pair* of chance-processes (one involving the
temporal evolution of the state of the primary system, that under observation,
and the other involving the temporal evolution of the state of our signal-
light). And we can start thinking in terms of a *single*, but two-stage, chance-
process, a single non-deterministic evolution occurring within a composite
physical system. The new evolving system is that composed of the primary
physical system under observation plus the signal-light. The two-stage
evolution of the composite system produces outcomes which are pairs of
the form (X,S) meaning that an outcome of the type X occurred in the primary
system, and this was followed by the green signal. It would be natural (and
formally possible) to refer to this pair using a notation like $X \cap S$; and
similarly to use S to refer to the composite outcome (Ω, S)—where Ω is the
set of all outcomes of the primary system. The symbol S then describes the
event 'the signal light came on', irrespective of what happened in the primary
system.

Once that is done, the formula for the posterior probability simplifies. Now
invoking nothing but probabilities of outcomes of a single (albeit composite)
chance-process, it declares the posterior probability to be the familiar ontic
conditional $P(X|S) = P(X \cap S)/P(S)$.

Of course, this looks just like the formula in the version of the Rule of
Conditionalization that has been the target of this whole discussion. But the
similarity is merely superficial, for the symbols have quite different meanings.
The big difference is that my formula incorporates the probabilities associated
with the observer's receipt of the information, while the Standard Formulation
ignores this. The latter's denominator $P(Y)$ is simply the probability of the
outcome Y being achieved in the primary sub-process within our composite. It
makes no essential reference to the means by which information about that
outcomes is transmitted. The denominator in my formula by contrast makes no
reference to the outcomes of the initial process. It refers exclusively to the
behaviour of the signalling process.

4.7 An illustrative example

The discussion of the last two sections was inconveniently abstract, and it is a good idea to see how it applies to concrete situations. So let us elaborate our discussion: by (firstly, in the present section) observing the signal-light analysis avoids the defective answer ($\frac{1}{3}$) to the two-boy family problem; and (secondly, in the next section) by using the abbreviated model alone to systematically evaluate one of the puzzling probabilities associated with a traditional paradox.

The observational-process we imagine in our first illustration is going to be convoluted, perhaps inconveniently so—but there are several good reasons for tolerating the inconvenience: firstly, because the convolutions model the fact that information-flow is often complex, with information sometimes withheld, and sometimes released, yet for different reasons in different circumstances, some of them erratic, and some of them systematic or deliberate; and secondly, to demonstrate the ease with which the 'signal-light modelling' can handle the computation. Artificial simplicity would obscure that fact.

So let us suppose that, after selecting the two-child family, a fair coin is tossed by the scrutineer. If it comes down heads, the scrutineer inspects the family. If it comes down tails, nothing else happens and no light comes on. When the inspection is carried out, the scrutineer tosses the coin again if the family is found to contain two boys, and turns the green light on, if she then gets a head. If however, the family only contains one boy, she also turns on the green light. If finally, the inspected family contains no boys, she turns the red light on.

Prior to seeing the signal-light, a rational observer would assigns a probability of $\frac{1}{4}$ to there being two boys in the family. Suppose this observer sees the green-light, but gets no other relevant information, beyond the details above as to how the light is operated. What updated probability should he then assign to there being two boys in the family?

To answer this question now, we only need to ascertain two things. The first of them is what I have designated $P(S)$ in Section **4.6** above, the probability the green signal-light comes on. This is just $\frac{1}{2} \cdot (\frac{1}{4} \cdot \frac{1}{2} + \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 0) = \frac{1}{2} \cdot (\frac{1}{8} + \frac{1}{2}) = \frac{1}{2} \cdot \frac{5}{8} = \frac{5}{16}$.

The second of them is $P(X \cap S)$, the probability that BOTH there are two boys in the family ['X'] AND the green signal-light comes on. This is just $\frac{1}{2} \cdot \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{16}$.

The posterior probability is then $P(X \cap S)/P(S) = \frac{1}{16} \div \frac{5}{16} = \frac{1}{5} = 20\%$.

In this case, the observer knows for certain that the family contains at least one boy if the green light comes on, but his newly acquired knowledge of the family is no more extensive than this. So if the scrutineer had refused to tell him the details behind the operation of the light (only assuring him that whenever the light was green there would definitely be a boy in the family),

the conditions for the application of the Standard Formulation would be met. If that formulation were valid, the posterior probability would be $\frac{1}{3}$.

So an observer who believed the Standard Formulation could be badly exploited. The scrutineer could, for example, offer to pay him $4 every time the family was found to contain two boys after the green light came on. For this privilege, the observer would pay the scrutineer $1 every time the green light came on. The observer would think he was sure to win, since for him (believing the Standard Formulation) the fair pay-out would be just below $3.34. But he would lose money fast, for he would only win on 20% of bets. He would thus get back an average $4 of every $5 he handed over. Yet it would never be true that he had been deceived by the scrutineer, and he would find no evidence (beyond his obvious losses) to support such an hypothesis: the family would truly be found to contain one boy every time she turned the green light on.

So the two rules are very different. The observer should refuse to play unless told how the scrutineer is deciding to turn on the green light. The Standard Formulation is seriously misleading about the fair bets. That is the core message of all the discussion above.

5 Some traditional-information puzzles: Monty Hall

A number of traditional problems in probability (like 'Bertrand's Box Paradox', 'The Two-Aces Puzzle', or 'The 3 Prisoners Puzzle', etc.)[23] derive their impact from Korb's 'propositional mesmerization' (p. 12), and the associated confusion (extricated above) between what I have designated Y and Y^*. They baffle us by displaying a conflict between several plausible revisions of a prior probability after new information is received. Yet once one accepts that a knowledge of *what* is discovered (Y) is rarely enough to allow a sound revision—and then pursues the missing Y^*—much of the puzzlement disappears, and the principles above provide a quick (and seemingly algorithmic) method of effecting the revisions, a method which enables us to avoid altogether the agonizing *ad hoc* manoeuvres (of uncertain validity) that plague discussion of these problems.

To illustrate this, let us conclude with a brief analysis of one of the traditional paradoxes, 'The Monty Hall Problem':

[23] For the box problem, see Kyburg ([1970], pp. 34–5), plus Schlesinger ([1991], pp. 188–90). For the two-aces puzzle, see Freund ([1965], pp. 29, 44), plus Gamov and Stern ([1958], pp. 37–42). For the puzzle of 3 prisoner's (*not* 'the [2] prisoner's dilemma'!), see Schlesinger ([1991], pp. 8–11), plus Jeffrey ([1988], pp. 226–8). For an extended discussion of the 'Monty Hall' debate, see Vos Savant and Granberg ([1997], pp. 5–17, 169–96), plus (for a discussion that contrasts sharply with mine) Howson and Urbach ([1993]), pp. 384–5, where the relevance of the compère's protocol is effectively denied. Discussions of the similar 'two-child family problem', have been cited frequently above, especiallly in fns 5, 6 and 7. For further citations of discussions of analogous problems, see Shafer ([1985], p. 264). For the quite different 2-prisoner dilemma, see e.g. Jeffrey ([1983], pp. 15–20).

A contestant in a game show ('Mary') is shown three opaque doors in a wall, one white, one black, and one red. She is told by the compère ('Monty') that a valuable prize has been placed (at random, for simplicity) behind one of these doors, but a worthless picture of a goat is behind each of the other two, and that she will get the prize if she can choose which door conceals it. Mary nominates the white door as her choice, to which Monty responds by opening the red door, revealing a goat. He then offers Mary the opportunity of switching her choice to black—in the light of the new information he has apparently just provided. What should Mary do?

To answer this question systematically, it is enough to produce a convincing evaluation of the probability Mary should allocate to the sentence W ($\equiv$ 'The prize is behind the white door)', after receipt of the new information provided by Monty. If this is greater than $\frac{1}{2}$, she should avoid changing to the black door. If it is equal to $\frac{1}{2}$, she neither gains nor loses by the change. If it is less than $\frac{1}{2}$, she should change. In the light of the discussion above, the probability we need to evaluate is just the ontic conditional $P(W|\bar{R}^*)$ where $\bar{R}^*$ is the sentence 'Monty reveals a goat behind the red door'. So we can readily evaluate the probability we seek using the Standard Formula $P(W \cap \bar{R}^*)/P(\bar{R}^*)$.

Yet to evaluate each of $P(W \cap \bar{R}^*)$ and $P(\bar{R}^*)$ we clearly need to penetrate Monty's psyche—a feature of the problem that sometimes seems puzzling. So *IF* Monty has just chosen randomly between the doors not selected by Mary (and thus run the risk of revealing the prize to her), then $P(\bar{R}^*) = \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot 0 = \frac{1}{3}$; while $P(W \cap \bar{R}^*) = \frac{1}{6}$, so the posterior probability becomes $\frac{1}{2}$, and the new information should not induce Mary to change. (But to realize this herself, Mary of course needs to know how Monty made his choice—in accordance with our more general observations in Section **4.3** above.)

But if Monty had decided only to reveal a door that was not hiding the prize, the situation is dramatically different. Suppose indeed he had decided that whenever Mary chose the white door, he would proceed as follows: if the prize is behind the white door, open the black one; but otherwise, open the door that does not hide the prize. Now, $P(W \cap \bar{R}^*) = 0$, so the posterior probability of the prize lying behind the white door is reduced to 0. Mary must change her choice—if she knows what Monty is doing. For in such circumstances, she would not have received her particular piece of new information if her initial choice had been a winner.

Mere calculations (like these) do not, of course, remove the psychological and logical barriers to understanding the Monty Hall problem, but those difficulties should have been significantly reduced by our earlier analysis. As puzzles like this have been much debated elsewhere, I will not attempt a broader analysis here. My main point is simply to suggest that these paradoxes are very much the product of a false conceptualization of the conditioning process; and, conversely, to indicate how easy it is to derive (and defend) some

of the confusing answers by applying my suggested revision of the misleading Rule.

Acknowledgements

Department of History and Philosophy of Science
University of Melbourne
Parkville, 3052, Australia
k.hutchison@hps.unimelb.edu.au

References

Bar-Hillel, M. and Falk, R. [1982]: 'Some Teasers Concerning Conditional Probabilities ', *Cognition*, **11**, pp. 109–22.

Billingsley, P. [1986]: *Probability and Measure*, second edn, New York: Wiley.

De Finetti, B. [1974]: *Theory of Probability: A Critical Introductory Treatment*, vol. 1, trans. A. Machí and A. Smith, London: John Wiley & Sons.

Dudewicz, E. and Mishra, S. [1988]: *Modern Mathematical Statistics*, New York: Wiley.

Falk, R. [1993]: *Understanding Probability and Statistics: A Book of Problems*, Wellesley, MA: Peters.

Feller. W. [1968]: *An Introduction to Probability Theory and its Applications*, third edn, rev. printing, vol. 1. New York: John Wiley & Sons.

Freund, J. [1965]: 'Puzzle or Paradox?', *American Statistician*, **19**, 4 (October), pp. 29, 44 (and ongoing discussion in later numbers).

Gamov, G. and Stern, M. [1958]: *Puzzle-math*, London: Macmillan.

Glymour, C. [1980]: *Theory and Evidence*, Princeton, NJ: Princeton University Press.

Grimmett, G. and Stirzaker, D. [1993]: *Probability and Random Processes*, second edn, Oxford: Clarendon Press.

Hodges, J. and Lehmann, E. [1964]: *Basic Concepts of Probability and Statistics*. San Francisco: Holden-Day.

Howson, C. and Urbach, P. [1993]: *Scientific Reasoning: The Bayesian Approach*, second edn, Chicago: Open Court.

Isaac, R. [1995]: *The Pleasures of Probability*, New York: Springer.

Jeffrey, R. [1983]: *The Logic of Decision*, second edn, Chicago: University of Chicago Press.

Jeffrey, R.[1988]: 'Conditioning, Kinematics, and Exchangeability', in B. Skyrms and

W. Harper *(eds)*, *Causation, Chance, and Credence*, vol. 1, Dordrecht: Kluwer, pp. 221–55.

Korb, K. [1994]: 'Infinitely Many resolutions of Hempel's Paradox', in R. Fagin *(ed.)*, *Theoretical Aspects of Reasoning about Knowledge*, San Francisco: M. Kaufmann, pp. 138–49.

Kyburg, H. [1970]: *Probability and Inductive Logic*, London: Macmillan.

Kyburg, H. [1990]: *Science and Reason*, Oxford: Oxford University Press.

Maher, P. [1993]: *Betting on Theories*, Cambridge: Cambridge University Press.

Nickerson, R. [1996]: 'Ambiguities and Unstated Assumptions in Probabilistic Reasoning', *Psychological Bulletin*, **120**, pp. 410–33.

Resnick, M. [1987]: *Choices: An Introduction to Decision Theory*, Minneapolis, MI: University of Minnesota Press.

Rosenkrantz, R. [1977]: *Inference, Method and Decision: Towards a Bayesian Philosophy of Science*, Dordrecht: Reidel.

Salmon, W. [1967]: *The Foundations of Scientific Inference*, Pittsburgh, PA: University of Pittsburgh Press.

Savage, L. [1954]: *The Foundations of Statistics*, New York: Wiley.

Shafer, G. [1985]: 'Conditional Probability', *International Statistical Review*, **53**, pp. 261–77.

Skyrms, B. [1986]: *Choice and Chance: An Introduction to Inductive Logic*, third edn, Belmont, CA: Wadsworth.

Van Fraassen, B. [1984]: 'Belief and the Will', *Journal of Philosophy*, **81**, pp. 235–56.

Vos Savant, M. (and Granberg, D.) [1997]: *The Power of Logical Thinking*, New York: St. Martin's Griffin. (Granberg is author of the appendix, pp. 169–96.)