Observations on Bayes' Theorem, logic and the admissibility of propensity evidence under the Evidence Act 2006[NZ][1]

Don Mathias
Barrister, Auckland

Mathematics and the logic of proof

One of the purposes of the Evidence Act 2006[NZ], heading a list of six in s 6, is "providing for facts to be established by the application of logical rules". This emphasis on logic sounds like a promise that the rules of evidence will be intrinsically reasonable or commonsensical. Unfortunately, there is a conflict between what common sense suggests, and what logic dictates.

This conflict may be seen from approaches to assessments of probabilities. Here are some examples:

(i) The other child
We are told by someone that the person over there with a child, a girl, has another child who is not present. All these people are strangers to us. What is the probability that the other child is also a girl?

(ii) The concealed prize
We are invited to guess in which one of three boxes, labeled A, B and C, there is a prize. Only one box has a prize in it, and after we have chosen box A, we are told that the prize is not in box C. We are then asked whether we want to change our choice. Should we?

You can tell from the context in which I have mentioned these examples that what seems to be the logical answers may not be correct. Indeed, virtually everyone thinks, wrongly, that there is a 50% chance that the other child is a girl,[2] and that there is no particular reason to change the choice of box.[3]

People even argue against the explanations of the correct answers to these examples, because we, as humans, are not particularly good at logic beyond a fairly elementary level.

---

[1] To check for updates to this draft paper, go to www.nzcriminallaw.blogspot.com and follow the link on the left to papers available at this site.
[2] See http://www.uiowa.edu/~030116/116/articles/mathrec1.htm for a transcript of an article by Ian Stewart in Scientific American, September 1996, p 134.
[3] http://www.americanscientist.org/template/BookReviewTypeDetail/assetid/15580;jsessionid=aaa5LVF0 is a review of Ian Stewart's book *The Magical Maze: Seeing the World through Mathematical Eyes*, (1998) in American Scientist September-October 1998; the review mentions this problem, which it calls the problem of the goats. For further discussion see http://www.willamette.edu/cla/math/articles/marilyn.htm . It is also known as the Monty Hall problem, and the solution is explained in a video available at http://www.youtube.com/watch?v=mhlc7peGlGg .

The logic of mathematics can be counter-intuitive. Consider the following:

(iii) The disease test
On average, one person in every 100,000 has a particular disease. There is a test for the disease, and the test is accurate 99% of the time. Does a positive test result mean the person tested is more likely than not to have the disease?

Simple arithmetic demonstrates that a positive test result here is most unlikely to mean the person has the disease.[4] One test result in 100 will be wrong, so there will be 1000 wrong results in every 100,000 people tested, and only one person will have the disease: the likelihood of the test result being correct is less than one in 1000. What the test has done, however, is change the chance of the person having the disease from one in 100,000 to one in slightly over 1000. The difference is between asking, as the problem asks, whether the person is "more likely than not" to have the disease, and whether the person is "more likely than before" to have it.

The points I make, from these three examples, are that people have difficulty in assigning probabilities, in recognising when a probability should be assigned to a class of events rather than to a particular event, and in perceiving the implications of probabilities.

It is proper to bear in mind, at this point, the fundamental differences in approach that have been taken to probability. To put it in general terms, with the inevitable sacrifice of accuracy, on the one hand are the mathematical theories which can be called Pascalian: they proceed on the basis that a probability of an event is a number between 0 and 1 which may be combined with one or more other probabilities by applying specific rules which derive from the logic of mathematics. Baysean logic is a Pascalian approach to probability. On the other hand, there are non-mathematical theories of probability, broadly termed Baconian, a recent prominent proponent of which has been L Jonathan Cohen.[5] Baconian probabilities are non-numerical but can be ordered, and reflect the weight that a proposition carries in its context, subject to rational argument. A convenient summary of the position is given by Dorothy Coleman:[6]

> "Cohen has proposed that what all conceptions of probability have in common is that they provide different criteria for grading degrees of

---

[4] Robertson and Vignaux, *Interpreting Evidence* (1995), p 19.
[5] *The Probable and the Provable* (1977).
[6] "Baconian Probability and Hume's Theory of Testimony" Hume Studies, vol 27 No 2, November 2001, p 195 at 198. A more detailed summary of the kinds of probability theories is given in the article by Philip Dawid, cited below, note 16. A different classification of the various theories of probability is given by L. Cosmides and J. Tooby, *"Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under certainty"* Cognition 58 (1996), 1, for example at p 4: "There are deep divisions among professional probabilists, not only over how statistical inference should be conducted, but over the very meaning of the word "probability". One of the deepest divisions is between the Bayesians and the frequentists." This comes from a supposition that, p 3: "Bayesians argue that probability refers to a subjective degree of confidence", whereas a better view is that Bayesean logic can use probabilities that are assessed subjectively or are based on observations of frequencies, and it is a bridge between Pascalian and Baconian approaches to probability.

> *provability*, and that degrees of provability allow for two kinds of scales. Pascalian scales take the lower extreme of probability to be disprovability or logical impossibility; the Baconian scale takes the lower extreme to be only non-provability or lack of proof. Because Baconian probability uses a different lower extreme than Pascalian probability, mathematical axioms that apply to the latter, such as complementational rules of negation, addition, and multiplication, do not apply to the former, with the consequence that degrees of Baconian probability are ordinal but not mathematical. Contextual considerations determine which scale is more useful or appropriate to employ. Baconian gradations of provability, its advocates argue, are particularly appropriate for assessing differential weight or relevance of evidence…." [Coleman's emphasis, footnotes omitted]

It seems that currently the courts apply a Baconian approach to propensity evidence, as indeed they do to evidence of most sorts; Baysean logic appears to be reserved, when it is accepted at all, for application to scientific expert opinion. Nevertheless, Pascalian probability theory is more popular among theorists, and Baysean logic has considerable support.

The question that should be borne in mind is whether any conclusion that is illogical when considered in Pascalian terms, would also be illogical when arrived at by Baconian reasoning.

Propensity and probability

Propensity evidence is like conditional probability. Given that something has happened, what is the probability that something else is true? Given the previous bad conduct, what is the probability of guilt on the present charge? That is the question that it seems natural to ask, but, as we shall see, it is not the logically correct one. An appropriate mathematical approach uses Bayes' Theorem.[7] The evidence of prior misconduct is incorporated into what is called a likelihood ratio to indicate its effect on the probability of guilt of the present offence that is raised by the other evidence in the case. Another word for likelihood ratio, more familiar to lawyers, is probative value. The likelihood ratio has, as its numerator, the probability of the occurrence of the prior misconduct, if the accused is guilty, and as its denominator, the probability of the occurrence of the prior misconduct, if the accused is innocent. If the likelihood ratio is near one, it has little effect on the prior probability of guilt, and it would be appropriate to say that the probative value of the propensity evidence was low, and, further, that it should be ruled inadmissible.[8]

---

[7] Above, note 4, especially at Chapter 2. See also the explanation of Bayes' Theorem that I have set out below, in the Appendix to this paper. I have assumed here that the prior event has (beyond reasonable doubt) happened, but this assumption is not necessary, as discussed below under the heading "Logic and s 43 of the Evidence Act 2006".

[8] Ibid, p 22.

Put in relevant mathematical terms, one way to formulate Bayes' Theorem is as follows:

$$\frac{P(\text{accused is guilty})}{P(\text{accused is innocent})} = \frac{P'(\text{accused is guilty})}{P'(\text{accused is innocent})} \times \frac{P(\text{propensity evidence} \mid \text{accused is guilty})}{P(\text{propensity evidence} \mid \text{accused is innocent})}$$

Here, P means "the probability of" the thing in brackets. The first fraction after the equals sign is called the prior probabilities, or the priors, and is indicated by the P′ signs. It may be thought of as the ratio of the probabilities yielded by the non-propensity evidence in the case. The numerator is the probability of the accused being guilty as indicated by the non-propensity evidence, and the denominator is the probability of the accused being innocent as indicated by the non-propensity evidence. The second fraction after the equals sign is the likelihood ratio. The numerator, top line, is a conditional probability, as is indicated by the vertical line. It means, the probability of finding the propensity evidence given that (ie, on the assumption that) the accused is guilty. The denominator of the likelihood ratio, the bottom line, is the probability of finding the propensity evidence given that the accused is innocent. This likelihood ratio is multiplied by the priors. The product of these two fractions on this side of the equals sign, the likelihood ratio and the priors, is equivalent to the left side of the equals sign: the ratio of the ultimate probability of the accused being guilty to the ultimate probability of the accused being innocent.[9]

Two points to note here are firstly that the probability of the accused being "innocent" is different from "not guilty" in the legal sense, and secondly, whether a verdict of guilty is appropriate will depend on what sort of value jurors require for the ratio on the left side of the equals sign. This is not to say that attempts should be made to put a mathematical probability on what is a reasonable doubt, but it has been observed that opinions vary widely about what sort of probability of innocence amounts to a reasonable doubt.[10]

I should mention here two other kinds of errors of logic that all too frequently accompanies reasoning about conditional probabilities:

---

[9] There are various other ways to express Bayes' Theorem. I have followed the one used by Robertson and Vignaux, above, because it is expressed appropriately for application to evidence in court. Numerous articles describing the theorem differently can be found online. For a brief defence of the application of the Baysean approach in court, see Robertson and Vignaux at http://www.mcs.vuw.ac.nz/~vignaux/docs/Adams_NLJ.html . Readers should not be discouraged by the impish remark of Dawid, below note 16 at p 61 of his Appendix: "Simple though it is, both the logic and the calculation involved in applying Bayes's theorem is likely to be beyond most judges (and even some jurors)."

[10] See the discussion in *R v Wanhalla* [2006] NZCA 229, especially the judgment of Glazebrook J on this point at paras 73-95.

Transposition of the conditional

This particularly technical-sounding name is a short way of saying that the things on either side of the vertical line are swapped around: it is wrongly assumed that $P(A|B) = P(B|A)$. An obvious illustration makes this easier to remember. If A is "this person is a male" and B is "this person is over 6 feet tall", then it is clear that the probability of a person being male if they are over 6 feet tall is not the same as the probability of their being over 6 feet tall if they are male.[11] The probabilities here are about samples from two different populations. The first is the population of people who are over 6 feet tall, and the second is the population of people who are male. More generally, $P(A|B)$ is a probability statement about the population having the characteristic B, and $P(B|A)$ is a probability statement about the population having the characteristic A.

The error of transposition of the conditional can occur in relation to the likelihood ratio. When this happens it is an example of "the prosecutor's fallacy". In respect of the numerator of the likelihood ratio, the error converts a statement about the population of guilty people into a statement about the population of people having the relevant characteristic. Here the characteristic is the propensity. For example, where the present charge is burglary, and the propensity evidence is a series of burglaries committed by the accused, the numerator is the probability of the occurrence of this series of burglaries on the accused's record assuming that the accused is guilty. This would reflect the rate of recidivism among burglars, and may be something like 0.75. The error of transposing the conditional here would involve saying that the probability that the accused is guilty, given that he has this record of burglaries, is 0.75. A statistic about the population of burglars cannot be taken to be a statistic about the population of burglars with this record of previous convictions. It may be that, after a series of convictions, burglars tend to decide that burglary isn't worthwhile. Another aspect of the error is that it ignores the other evidence in the case, the priors. The probability of guilt (the numerator on the left side of Bayes' Theorem) requires taking more into account than just the numerator of the likelihood ratio.

Another prosecutor's fallacy involves the denominator of the likelihood ratio.[12] The denominator cannot be transposed into a probability of innocence. P(propensity|innocence) is not equivalent to P(innocence|propensity). The sampling of the population of innocent people is not the same as the sampling of the population of people with the relevant propensity. A very low likelihood of the occurrence of the propensity in innocent people is not the same as a very low likelihood of innocence on the present charge. Again, this form of the error of logic involves ignoring much of the formula in Bayes' Theorem.

[11] This illustration of transposing the conditional is given in Robertson and Vignaux, above, note 4, p 20.
[12] Robertson and Vignaux, above, note 4, pp 92-93.

The misuse of coincidence

It is not unusual to hear prosecutors say to juries that one piece of evidence against the accused may have an innocent explanation, but together with another piece of evidence, both can't be just coincidence, their joint occurrence must indicate guilt. This argument may be made in relation to any sort of evidence, not just evidence of propensity. A more generalised form of a coincidence argument, and one that can apply at a preliminary stage, is that the joint occurrence of two items of evidence increases the likelihood of both of them being true.

For example, suppose the issue at trial is whether the accused was a robber. He was found, some days after the crime, to be in possession of the same kind of jersey that an eyewitness described the robber wearing. Also, he had possession of the same kind of balaclava as the robber was seen wearing. The first item of evidence is used in support of the prosecutor's assertion that the jersey found in the accused's possession is the same jersey that the robber wore. The second item of evidence is used in support of the assertion that the balaclava found in the accused's possession is the same one worn by the robber. In formal language appropriate to probabilities, event A is "the jersey is the one the robber wore", and event B is "the balaclava is the one the robber wore". P(A) is the probability that the jersey is the one that the robber wore, and P(B) is the probability that the balaclava is the one that the robber wore.

At the preliminary stage of deciding what P(A) and P(B) are, the prosecutor may suggest that the coincidence of the joint occurrence of A and B increases the probability that each is true. Indeed, it does, but there is a risk of exaggerating the effect of the coincidence on the probabilities of A and B. In any event, it is not necessary for the probability of the coincidence of the joint occurrence of A and B to be worked out. This is because Bayes' Theorem is a consequence of the joint occurrence of the items of evidence, and adding coincidence as a separate consideration would be superfluous. The probability of coincidence is P(A and B), which may be written $P(A \cap B)$, and this is not a term in Bayes' Theorem.[13] Nevertheless, for the sake of illustrating another potential error, I set out here how the coincidence assessment should be done. The Baysean probability of the joint occurrence of A and B is P(B|A) x P(A), or, equivalently, P(A|B) x P(B).[14] That means, for example, asking what is the probability of the balaclava being the one the robber wore on the assumption that the jersey had been worn by the robber, and then multiplying that by the probability of the jersey having been worn by the robber. That probability of a joint occurrence of the finding of the jersey and the balaclava indicates a revision of the individual probabilities which were based on the assumption that A and B were independent. For example, if on their own (ie, independently), P(A) = P(B) = 0.4, and if P(B|A) was assessed at, say, 0.9, then the probability of the coincidence is 0.36, and the revised P(A) = P(B) = 0.6 if A and B were independent. This is more than the

---

[13] It is a term in the derivation of the theorem: see http://en.wikipedia.org/wiki/Bayes'_theorem . Reversion to $P(A \cap B)$ in argument may be thought of as ignoring the implications of the coincidence and inviting duplication of the role of coincidence. In fact, $P(A \cap B)$ = P(A|B) x P(B) = P(B|A) x P(A), where A and B are not independent, and where they are, P(A and B) = P(A) x P(B).

[14] See Appendix.

initial assessment of each, but it is less than the conditional probability P(B|A). That conditional probability is the exaggeration, if it is wrongly regarded as the last step in the assessment, that must be avoided: if one were to ask what the effect of coincidence was, it would be necessary to complete the reasoning by ascertaining the revised probabilities.

As indicated, Bayes' Theorem does take coincidence into account. The theorem emerges from the implications of $P(A \cap B)$, so if the theorem is applied to the reasoning process an additional resort to the unlikelihood of coincidence would be superfluous. This danger is particularly acute when an expert witness has given evidence in the form of a likelihood ratio. It would be wrong for a prosecutor to invite the jury to consider the unlikelihood of coincidence on top of that.

At the outset, it is necessary to assess what values to give P(A) and P(B). Those will depend on the context of the evidence, including subjective factors like how reliable the witnesses seem to be. Bayes' Theorem shows how these items of evidence combine to affect the probability that the accused is guilty. A likelihood ratio can be constructed for each item, A and B, and the priors are multiplied by these likelihood ratios, in this expanded form of the theorem:[15]

$$\frac{P(G)}{P(NG)} = \frac{P'(G)}{P'(NG)} \times \frac{P(A|G)}{P(A|NG)} \times \frac{P(B|G)}{P(B|NG)}$$

The first likelihood ratio, which I have highlighted, is the probability of the jersey having been that worn by the robber given that the accused is guilty (this would be the same as P(A) as assessed above) divided by the probability of the jersey having been worn by the robber given that the accused is innocent (this would be low, but might be higher if the robber may have given it to the accused, or if the accused may have lent it to the robber). If the likelihood ratios are greater than 1 they will increase the ratio of the probability of guilt to the probability of innocence. For example, if P(A) = P(A|G) = 0.4, and if P(A|NG) was assessed as low, say 0.2, then this likelihood ratio is 2: it doubles the priors. If the likelihood ratio involving B was also 2, the effect would be another doubling of the priors. That is, the two items of evidence would together multiply the priors by 4.

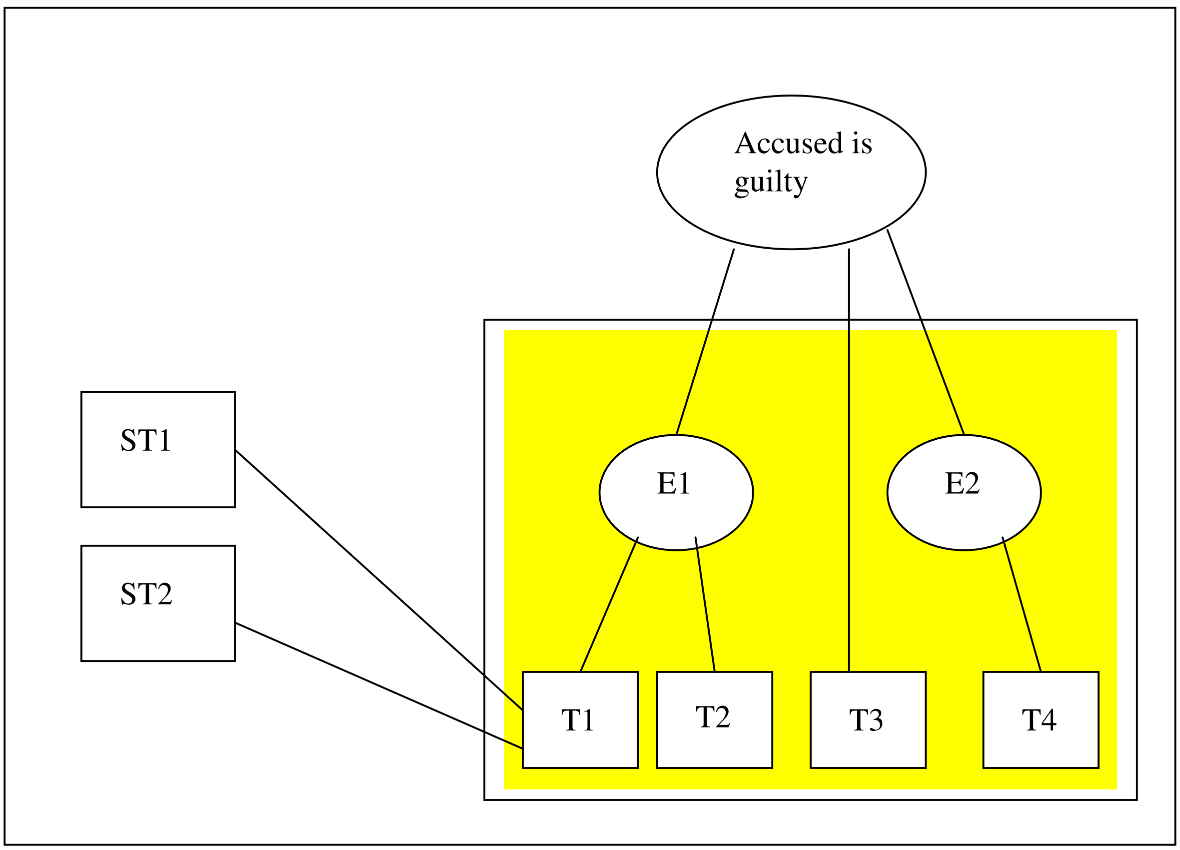
The relevance of this to propensity evidence is readily apparent. Event A may be "the complainant is telling the truth about what the accused did" and event B may be "the accused has done this sort of thing before". The probability of event A will properly be taken into account in assessing the priors, but once singled out by the prosecutor its place in the reasoning is as shown in the expanded form of Bayes' Theorem above. The prosecutor argues that these events can't be coincidence, and the defence (as well as the

[15] This formulation assumes A and B are independent; where they are not, A must become a second condition for the ratio involving B. See Robertson and Vignaux, *Interpreting Evidence* (1995) at pp 227 – 228. This level of precision is unlikely to be necessary for fact-determiners in law, it is the avoidance of logical error that is important.

judge) has to ensure the jury does not wrongly add some sort of weighing for coincidence on top of a likelihood ratio.[16]

Using propensity evidence

The mathematical model is useful to indicate when propensity evidence should be admissible, and to identify some logical fallacies, but it does not show how items of propensity evidence should be used in the light of other evidence in the case. How does the probative value of the propensity evidence interact with other evidence? In the diagram below, I have separated two items of propensity evidence, ST1 and ST2, from the other evidence in the case (in the yellow box). This other evidence is the evidence of witnesses T1 – T4, and inferences from that testimony, E1 and E2. The diagram illustrates how the similar fact evidence may influence the probative value of an item of testimonial evidence. The lines indicate the paths of such influence in an imaginary case.



[16] See further, the particularly illuminating discussion, "Probability and Proof" by Philip Dawid, which is his Appendix to "Analysis of Evidence" by T.J. Anderson, D.A. Schum and W.L. Twining: http://www.cambridge.org/resources/052167316X/2870_reformatted%20Appendix%20I%20for%20website.doc . His discussion, at pp 66 – 69, of the English case *R v Adams* illustrates the difficulties judges may have in accepting mathematical constraints to the jury's approach to reasoning.

This is not the only way propensity evidence may operate. It may have a bearing on the weight to be given to an inference, E1, drawn from testimony. There would be a risk of double-counting, or circular reasoning, if ST1 were thought to bear upon both T1 and E1, instead of bearing on E1 through T1, or on E1 independently of T1.

I have not attempted here to make use of two kinds of diagrammatic representations of legal reasoning, called Wigmore charts and Baysean networks respectively, because my point is general and not an attempt to analyse the facts of a particular case, which is what those methods are designed to do.

It follows from Bayes' Theorem and from the diagram above, that the likelihood ratio generated by the propensity evidence in relation to the hypothesis that the accused is guilty is the same as the likelihood ratio generated by that evidence in relation to the testimony to which it is relevant, T1. That is,

$$\frac{P(ST1|G)}{P(ST1|NG)} = \frac{P(ST1|T1)}{P(ST1|NT1)}$$

And, focusing on the fact-finder's task of evaluating T1 in the light of ST1:

$$\frac{P(T1)}{P(NT1)} = \frac{P'(T1)}{P'(NT1)} \times \frac{P(ST1|T1)}{P(ST1|NT1)}$$

The likelihood ratio of the propensity evidence will be considered with the likelihood ratios for the other items of evidence; Bayes' Theorem accommodates this by the assessment of the priors, but the diagram above illustrates how items may interact. For example, item T4 may be defence evidence supporting an inference of innocence, E2; the likelihood ratio appropriate to this evidence will influence the outcome in a way that diminishes the influence of the likelihood ratio of the similar fact evidence. Another item may carry the day, for example if T3 is evidence of an irrefutable alibi (such as, the accused was in prison at the time of the alleged offence, which did not occur in prison); in such a case the numerator of the likelihood ratio would be 0, and P(G) = 0. In this sense, the way it interacts with other evidence, propensity evidence is like any other item of circumstantial evidence.

Logic and s 43 of the Evidence Act 2006

The Evidence Act 2006 does not state propositions concerning the probabilistic approach to propensity evidence. Nevertheless, the above considerations support the following. When determining the admissibility of propensity evidence, its probative value (the likelihood ratio) must be assessed without reference to the probative value of other evidence in the case. It would be wrong for the judge to say, "well, there is evidence suggesting the accused is probably guilty, so I'll let in the propensity evidence." This is because the denominator of the likelihood ratio is based on the *assumption* that the accused is innocent (the probability of the propensity evidence occurring *given that* the accused is innocent), and to lessen the denominator (and, thereby, to increase the likelihood ratio) by saying, the chance of innocence is low, therefore the probability of the occurrence of the propensity evidence if he is innocent is low, is a fallacy.[17]

Another point that emerges from the logic of propensity evidence is that its admissibility is appropriately dealt with before trial. It would be wrong for a judge to say, on a pre-trial ruling, that the admissibility of the propensity evidence should be left for the judge at trial, when a better perspective of the whole case will be obtained. Such a perspective is irrelevant, for the reasons given in the last paragraph.

This is not to suggest that there is no threshold of evidence of guilt that must be crossed before propensity evidence will be admissible. The function of the likelihood ratio is to operate on the prior probabilities: the ratio of the probability of guilt on the basis of the other evidence in the case, to the probability of innocence on the basis of that other evidence. If there is no other evidence of guilt, apart from the propensity evidence, the propensity evidence cannot have any effect on the probability of guilt. The likelihood ratio multiplied by zero prior probability of guilt is zero. This prevents the "round up the usual suspects" approach to proving guilt. On the other hand, the prior probabilities need not be particularly high before propensity evidence can be used. Even the balance of probabilities in favour of guilt is not necessary. As long as there is some evidence of guilt, the probative value of the propensity evidence will have something to operate on. But again, this is not to say that *any* probative value will be sufficient to render the propensity evidence admissible. Nor is it to say that, just because there is an assessable prior probability of guilt, the evidence of propensity *must* be admissible, as there are statutory criteria to be met first. The fundamental principle, contained in s 7 of the Act, that the primary requirement for evidence to be admissible is that it must be relevant, that is, it must tend to prove or disprove anything that is of consequence to the determination of the proceeding, amounts to a requirement that the likelihood ratio it generates must not be one, and it must vary from one to an extent that it has an appreciable effect on the prior probabilities.

A potentially difficult question is the extent to which the court, and ultimately the jury, needs to be satisfied that the propensity evidence is itself true. At common law in Australia it seems that the propensity evidence must be proved beyond reasonable doubt. That requirement could be seen as a quid pro quo for allowing propensity evidence at all.[18] But Baysean logic does not require that high standard. If, for example, the probability that the events disclosing the alleged propensity actually happened is only 0.8, that can be taken into account in assessing the likelihood ratio, and it will increase the

---

[17] In *R v Holtz* [2003] 1 NZLR 667, (2002) 20 CRNZ 14 (CA) the court observed, at para 38, that propensity evidence is just another form of circumstantial evidence, to be considered together with all the other evidence in the case, and in respect of which no particular standard of proof is required.

[18] *HML v R* [2008] HCA 16 (24 April 2008). The expression quid pro quo was used by Kirby J at 83. In New Zealand it has been held that similar facts do not have to be proved beyond reasonable doubt: *R v Guy* (1996) 13 CRNZ 589 (CA).

denominator and thereby reduce the effect of the propensity evidence on the priors. That would be consistent with common sense.

Probative value

Propensity evidence may be offered by a defendant about himself, by a defendant about a co-defendant, or by the prosecution. The provision of the Evidence Act 2006 that I will discuss here is s 43 which concerns propensity evidence offered by the prosecution. The use of the expression "probative value" in this section will be examined in the light of the above discussion of the logic of proof.

> **"43 Propensity evidence offered by prosecution about defendants**
>
> (1) The prosecution may offer propensity evidence about a defendant in a criminal proceeding only if the evidence has a probative value in relation to an issue in dispute in the proceeding which outweighs the risk that the evidence may have an unfairly prejudicial effect on the defendant.
>
> (2) When assessing the probative value of propensity evidence, the Judge must take into account the nature of the issue in dispute.
>
> (3) When assessing the probative value of propensity evidence, the Judge may consider, among other matters, the following:
>
> (a) the frequency with which the acts, omissions, events, or circumstances which are the subject of the evidence have occurred:
>
> (b) the connection in time between the acts, omissions, events, or circumstances which are the subject of the evidence and the acts, omissions, events, or circumstances which constitute the offence for which the defendant is being tried:
>
> (c) the extent of the similarity between the acts, omissions, events, or circumstances which are the subject of the evidence and the acts, omissions, events, or circumstances which constitute the offence for which the defendant is being tried:
>
> (d) the number of persons making allegations against the defendant that are the same as, or are similar to, the subject of the offence for which the defendant is being tried:
>
> (e) whether the allegations described in paragraph (d) may be the result of collusion or suggestibility:
>
> (f) the extent to which the acts, omissions, events, or circumstances which are the subject of the evidence and the acts, omissions, events, or circumstances which constitute the offence for which the defendant is being tried are unusual.
>
> (4) When assessing the prejudicial effect of evidence on the defendant, the Judge must consider, among any other matters,---

(a) whether the evidence is likely to unfairly predispose the fact-finder against the defendant; and

(b) whether the fact-finder will tend to give disproportionate weight in reaching a verdict to evidence of other acts or omissions."

The first point to note is that the weighing of probative value against the risk of unfair prejudice is not a requirement of the logic of proof as expressed in Bayesian analysis. The need to avoid unfair prejudice is not a requirement of logic, but of legal policy. It is a reflection of the absolute nature of the accused's right to a fair trial.[19] The way prejudice may be taken into account is discussed below under the heading "Assessing improper prejudice".

How do the matters listed in subsection (3) fit with the logic of proof? They are proper considerations because they will operate with logical consistency. For example, the first-mentioned matter is the frequency with which the acts (etc) have occurred. Where the present allegation is of misconduct that may be driven by a compulsion, a greater frequency of previous incidents may tend to increase the numerator of the likelihood ratio: that is, the probability of a frequent occurrence of previous misconduct may be higher if the accused is guilty. Similarly, if the accused is innocent it may be less likely that a frequent occurrence of previous misconduct would be found, so the denominator of the likelihood ratio may be smaller. Both these possibilities – higher numerator and lower denominator - tend to increase the likelihood ratio, and so are logically consistent.

Assessing improper prejudice

Evidence of propensity may be improperly prejudicial in various ways. It may distort the assessment of the likelihood ratio, by causing an inaccurate assessment of the numerator, the denominator, or both. It may be used wrongly to directly affect the assessment of the prior probabilities, and then, in an exercise akin to double counting, operate again on the priors via the medium of the likelihood ratio.

It is possible to read these mechanisms as examples of the matters referred to in s 43(4). Unfairly disposing the fact finder against the defendant is mentioned in para (a) of this subsection. This could include the use of the propensity evidence in the double counting way just mentioned, causing an inappropriate exaggeration of the prior evidence of guilt before the likelihood ratio is applied to it.

The other sort of prejudicial effect referred to is in s 43(4)(b): giving disproportionate weight to the propensity evidence. This could arise in the form of a distorted likelihood ratio, with an improperly increased numerator and/or an improperly reduced denominator.

---

[19] See the reference to the Bill of Rights in s 6 of the Evidence Act 2006, and the dicta on the accused's right to a fair trial in *Condon v R* [2006] NZSC 62 at para 77.

In legal terms, the court must guard against an error, in admitting propensity evidence, that would give rise to a real risk of loss of a more favourable verdict for the accused. This standard is used in the law concerning the avoidance of a substantial miscarriage of justice.[20] It is not a flexible standard: the accused does not have to accept a greater risk of loss of a more favourable verdict if wider policy considerations seem to support that increased risk.

On this approach, it can be seen that the weighing exercise referred to in s 43(1) is not an accurate metaphor.[21] It should not be read as requiring the accused to accept a larger risk of unfair prejudice in cases where the propensity evidence has higher probative value. To take that approach would be to tolerate trials that are unfair to the accused where the prosecution evidence is of high probative value, and that would be contrary to s 6(b) of the Act. The weighing exercise should be seen as an assessment of the risk of improper use of the evidence: probative value is akin to the likelihood that the fact-finder will use the evidence properly, and unfairly prejudicial effect is akin to the risk that the fact-finder will use the evidence improperly. Thus, if the risk of improper use can be overcome by an appropriate warning to the jury, the judge can conclude that the weighing exercise does not require exclusion of the evidence.

I should stress that I am not suggesting that jurors should be instructed on Bayes' Theorem and its formula for conditional probability calculations.[22] While Bayesian analysis is of particular use to witnesses who give expert scientific evidence, where data is readily available in probability form, the method is applicable to all forms of evidence. This is especially so in relation to propensity evidence, which is readily separable from the other evidence in the case because it concerns separate, or discrete, activity from that alleged in the charge presently under consideration by the court.

Propensity and identity

There is judicial support for the view that, when evidence of propensity is used to prove the identity of the present offender, a special approach is needed. In such cases, the present accused's own known propensity is sought to be used as proof that it was he who committed the present offence. The more general use of propensity evidence is to use the accused's own propensity as proof that he did do what it is now alleged he did. It is obvious that there is little difference between these uses of propensity evidence, and there is no obvious logical reason for a special approach to be required just because the present issue is the identity of the offender.

Three possible approaches where the present issue is the identity of the offender were reviewed in *R v Gee* [2001] 3 NZLR 729, (2001) 19 CRNZ 102 (CA): a sequential

[20] *Sungsuwan v R* [2006] 1 NZLR 730, (2005) 21 CRNZ 977 (SC), discussed in Mathias "Proof, fairness and the proviso" [2006] NZLJ 156.
[21] I have also criticised it in "Probative value, illegitimate prejudice and the accused's right to a fair trial" (2005) 29 Crim LJ 8.
[22] Stewart, above, note 2, makes the same point in his concluding paragraph: it is the avoidance of errors in logical reasoning that is important.

approach, a pooling approach, and a global approach. The sequential approach requires the jury to be sure that the accused committed a previous offence before it can be used as propensity evidence on the present charge. The pooling approach requires the jury to be sure, first, that the previous series of offending was committed by the same offender (whoever that might be) before then using all the evidence to decide whether that offender was the accused. The global approach rejects any suggestion of proof to any standard of the propensity evidence before it can be used, instead, it regards propensity evidence as circumstantial evidence, the inference of guilt being permitted if innocence would require too great a coincidence.

The logic of conditional probability does not require a special approach in cases where the offender's identity is in issue. In Bayes' Theorem the likelihood ratio involves conditionals that refer to the accused: the conditional in the numerator is that the accused is guilty, and the conditional in the denominator is that the accused is innocent. It is the accused's propensity that is relevant, and it is the likelihood of finding this propensity on the part of the accused that provides the other terms in the likelihood ratio. The unstated assumption is that it is possible to assess the probability that the propensity evidence is about the accused. But, in the *Gee* classification, a standard of proof is not specified: the pooling approach, and the sequential approach (which is really a special case of the pooling approach) require that the jury must be "sure" that the accused had the propensity, before that propensity becomes evidence on the present charge. In contrast, the global approach proceeds on the basis that no particular standard of proof is applicable, so that, in Baysean terms, a likelihood ratio is not used. In this respect the global approach is too lax, because a probability for the truth of the alleged propensity evidence should be assessed, whereas the pooling and the sequential approaches are too strict.

In *R v Holtz*[23] the Court of Appeal departed from the tripartite classification of approaches in *Gee*, holding that the basic requirement is that the probative value of the evidence outweighs illegitimate prejudice to the accused. This is now reflected in s 43(1) of the Evidence Act 2006. At the same time, the Court disparaged any attempt to find other generally applicable principles, and held that no special approach is required where the issue is identity. It held that, since the elements of the present charge have to be proved beyond reasonable doubt, where identity is in issue and the propensity evidence is the only, or substantially the only, evidence of identity, it must be proved to that standard. Crucially, the Court held that no particular standard of proof is applicable to the admissibility of propensity evidence, it being sufficient to tell juries that they must "find", "conclude" or "be satisfied" that the pattern or link is proved.[24]

In dispensing with a need for assessment of a probability that the propensity evidence is relevant to the accused, and in rejecting the search for generally applicable principles – here we are considering the relevance of conditional probability logic and the implications of Bayes' Theorem – the Court in *Holtz* has strayed from logical coherence. At the very least, the issue of the relevance of Bayesian analysis, and its lessons in logic,

---

[23] Above, note 17.
[24] See *Holtz*, above, note 17, at paras 35 – 39.

to the admissibility and use of propensity evidence, is open for further exploration. Expert evidence from mathematicians may be needed if logical rules on this subject are to be formulated.

Should we be Pascalians or Baconians?

I mentioned above that there is a question to be considered: are errors of Pascalian logic also errors of Baconian logic? Does it really matter that most people answer wrongly the three problems I set out at the beginning? Can a positive test result be given a probative weight that ignores a high incidence of false positives?

An interesting clash of opinions over whether fact-finders should be required to reason in a Pascalian way, as opposed to a more instinctive Baconian way, occurred in the English journal, the Criminal Law Review, in 1979 and 1980, in exchanges between Sir Richard Eggleston,[25] the Pascalian, and L Jonathan Cohen,[26] the Baconian. In a letter to that journal replying to an article by Sir Richard, Mr Cohen deals with the likelihood of the plaintiff succeeding in a hypothetical civil case where a person selected at random from the audience at a rodeo is sued for attending without paying:[27]

> "Suppose that in fact 85 people had climbed over the wall at the rodeo for every 15 who had paid for admission. On Sir Richard's view this would establish a very strong prima facie case that the randomly selected defendant had not paid: there is a Pascalian probability of seventeen-twentieths in favour of the plaintiff. But now suppose that an independent witness, of unimpeachable integrity, was standing by the automatic turnstile throughout the period in question and testifies to having seen the defendant pay, while it is agreed by both sides, as a result of thorough tests, that the witness's testimony to having seen or not seen a particular face in such a situation is 80 percent reliable because of his excellent eyesight and memory. *Surely most courts would take this as grounds for rejecting the plaintiff's suit.* Yet such a decision runs counter to Sir Richard's principles. If an initial Pascalian probability of seventeen-twentieths in favour of the plaintiff is amalgamated with a later Pascalian probability of four-fifths in favour of the defendant, we get a final figure of seventeen-twenty-ninths in favour of the plaintiff. (In the long run of such cases the witness would make 68 percent – four-fifths x 85 percent – correct identifications of people as having not paid, 12 percent correct identifications as having paid, 3 percent incorrect identifications as having

[25] Chancellor, Monash University, author of "Evidence, Proof and Probability"; Sir Richard died in 1991.
[26] Queen's College, Oxford, author of "The Probable and the Provable"; Mr Cohen died in 2006.
[27] [1980] Crim LR 748. I have omitted a footnote from this extract, and have changed Mr Cohen's use of the expression "per cent." to "percent". In addition to failing to appreciate the conclusion that properly follows from his own reasoning, Mr Cohen did not accurately assess the Pascalian position on the example he gives in his letter that immediately follows that quoted here.

> not paid and 17 percent incorrect identifications as having paid.)" [emphasis added]

Mr Cohen's calculations are correct, but he fails to emphasise that for every 29 occasions that the witness says a person paid, he is right only 12 times. Being correct only 12 out of 29 times is not strong evidence in favour of his assertion, as the witness is more likely to be wrong when he says a person paid. Yet Mr Cohen's Baconian position, above, is that "Surely most courts would take this as grounds for rejecting the plaintiff's suit." Perhaps not all Baconians would agree with Mr Cohen once they analyse the implications of the witness being 80% "reliable".

The Baysean formula, if applied here, would use as the priors 0.85 over 0.15, and the likelihood ratio generated by the witness's evidence would be 0.17 over 0.12, so the probability of guilt comes out as 0.88. This means the witness's evidence (that the defendant paid) has slightly increased the prior probability of the defendant not having paid from 0.85. Mr Cohen is right to say that the Pascalian and Baconian approaches have different results: the Pascalian result is counter-intuitive for at least some Baconians.

Plainly, both cannot be right. People who might have been inclined to take a Baconian approach could, on recognising Mr Cohen's error concerning the significance of the witness's reliability, be more sympathetic to a Pascalian stance. This requires attention to the implications of all the information upon which a solution to a problem depends. Mr Cohen's discussion of the above hypothetical shows how a Baconian may not recognise the full implications of a Pascalian approach, and a full Pascalian analysis reveals an error in the Baconian approach.

Appendix

A version of Bayes' Theorem for lawyers

Probabilities

The probability of an event E is expressed as P(E) and is 0 if E is impossible and 1 if E is certain to happen. Some probabilities are assessed by studies of the frequency of occurrence of E, and others are estimated against the experience of the person who assesses the likelihood of E.

The probability of two events happening is the product of the probability of each of them. For example, the probability of tossing a coin twice and getting two heads is 0.5 x 0.5 = 0.25, as can be seen from the possible outcomes HH, HT, TH, TT. The probability of occurrence of two events E and A is written as $P(E \cap A)$.

Conditional probabilities

What is the probability that E will happen, given that event A has happened? This question is expressed as what is $P(E|A)$. If A and E are independent of each other in the sense that the occurrence of one does not affect the likelihood of the occurrence of the other, then $P(A|E) = P(A)$, and $P(E|A) = P(E)$. Events in law are often not independent, especially if they are each brought about by the accused's guilt.

If event A is certain, $P(A) = 1$, so $P(E \cap A) = P(E|A)$.

Where $P(A)$ is less than 1, $P(E \cap A) = P(E|A) \times P(A)$.

If the conditional event is E, instead of A, the probability of both E and A occurring is $P(E \cap A) = P(A|E) \times P(E)$. These two ways of expressing $P(E \cap A)$ are equivalent, and can be written

$$P(E|A) \times P(A) = P(A|E) \times P(E)$$

Which is the same, dividing each side of the equation by $P(A)$, as

$$P(E|A) = \frac{P(A|E) \times P(E)}{P(A)}$$

This is a commonly-expressed version of Bayes' Theorem.

Adapting Bayes' Theorem for law

A variation is used in law, and it can be seen to emerge as follows. Suppose event E is "the accused is guilty" and event A is an item of evidence presented in the prosecution case. It is useful to compare the probability of guilt to the probability of innocence (or, in terminology used here, "not E"), so each side of the equation is divided by P(not E):

$$\frac{P(E|A)}{P(\text{not } E)} = \frac{P(A|E) \times P(E)}{P(\text{not } E) \times P(A)}$$

The terms on the right can be rearranged slightly:

$$\frac{P(E|A)}{P(\text{not } E)} = \frac{P(E)}{P(\text{not } E)} \times \frac{P(A|E)}{P(A)}$$

Putting these into English, the first term, the one on the left of the equals sign, is the ratio of the probability of guilt, in the light of the evidence A, to the probability of innocence. This ratio is sometimes called the posterior, as it refers to the position after event A is taken into account. It can also be called the odds of guilt given event A. The next term, after the equals sign, is the ratio of the probability of guilt, in the absence of the evidence A, to the probability of innocence. This ratio is called the priors, as it refers to the position before event A is taken into account. It can also be called the odds of guilt without evidence A. The third ratio in the formula is called the likelihood ratio: the ratio of the probability of finding the evidence A, given that the accused is guilty, to the probability of finding the evidence without that conditional. This formulation of Bayes' Theorem is called Bayes Theorem in terms of odds and likelihood ratio.[28]

[28] This terminology is used at http://en.wikipedia.org/wiki/Bayes'_theorem , where other ways of expressing the theorem can also be seen.