Probability Arguments in Criminal Law Illustrated by the Case of Lucia de Berk

Herman Philipse*

1. Prelude

Courses in Legal Reasoning offered by law faculties typically restrict the topic to forms of argument used in the application of legal rules. Such courses will comprehend methods of the interpretation of laws or precedents, rules for their application to particular cases, and some types of reasoning that we also find in other fields, such as arguments by analogy or *Fallvergleich*. However, if law faculties intend to train students in all patterns of reasoning that lawyers will have to practise in their professional career, they should widen the scope of these courses.

Must not instruction in legal reasoning include the many types of argument used in legislative drafting, for example? And what about the inquisitorial task of Continental criminal judges to establish the facts of the case before applying criminal law, as specified, for example, by the Dutch Code of Criminal Procedure (*Wetboek van Strafvordering*), Article 338 et seq.? Judges cannot entirely leave this task to scientific experts, since they have to understand and evaluate the experts' arguments. In this article, I focus on one kind of reasoning that is central to this latter endeavour: probability arguments concerning factual hypotheses. With regard to most criminal cases, the evidence only makes it probable to a certain extent that the defendant is, or is not, guilty. It follows that probability arguments nearly always play a crucial role. Do judges have a sufficient grasp of the logic of probability?

I shall start (Section 2) by reminding the reader of a(n) (in)famous Dutch criminal lawsuit, the case of Lucia de Berk. It is no accident that the final retrial of this case was triggered mainly by the work of a Dutch philosopher of science, and not primarily by lawyers. One of the eight pillars of De Berk's conviction was the so-called 'Coincidence Argument'. In Sections 3 and 4 the logic of this argument will be spelled out, and the question will be raised whether one can draw a conclusion concerning the *probability* of a hypothesis from a comparison of *likelihoods* (in this context, these are technical terms, the meaning of which will be explained in Section 3).

Paying insufficient attention to the logic of probability often results in fallacies, the most recurrent of which has been dubbed on good grounds the 'Prosecutor's Fallacy'. In Section 4, it will be argued that Bayes' Theorem should structure demonstrations that a defendant is guilty. Many experts have recommended Bayesianism with regard to cases such as Lucia de Berk's. Using this well-known case as an illustration, I argue that a Bayesian approach triggers some new questions, at least one of which I

* Prof. Dr. Herman Philipse, Distinguished Professor of Philosophy, Utrecht University, the Netherlands. E-mail: H.Philipse@uu.nl.

have not found in the existing literature. I conclude that a more comprehensive course on legal reasoning should be obligatory for law students (Section 5).[1]

2. A famous Dutch legal error

During the working hours of the nurse Lucia de Berk in the Juliana Children's Hospital (*Juliana Kinderziekenhuis*), The Hague, in 1999-2001, relatively many deaths (6) and reanimations of patients (5) had occurred, compared to the times during which the other nurses were on duty. Her past performance was investigated and the administration of the hospital reported the case to the Public Prosecutor. On 18 June 2004, the Court of Appeal of The Hague (*Gerechtshof 's-Gravenhage*) sentenced Ms de Berk to life imprisonment plus hospital detention and compulsory psychiatric treatment for seven cases of murder and three of attempted murder in the three hospitals in which she had worked.[2]

An appeal to the Dutch Supreme Court (*Hoge Raad der Nederlanden*) was largely rejected on 14 March 2006, after which Lucia suffered a cerebral infarct.[3] On 13 July 2006, the Amsterdam Court of Appeal (*Gerechtshof Amsterdam*) confirmed the sentence of life imprisonment.[4] An investigation of the case by the Dutch philosopher of science Ton Derksen, supported by his sister, the physician Metta de Noo-Derksen, critical discussions in the media, and a petition signed by many professors of statistics and the Nobel Prize-winning Dutch physicist Gerard 't Hooft, ultimately led to an exceptional retrial and to a complete acquittal by the Arnhem Court of Appeal (*Gerechtshof Arnhem*) on 14 April 2010.[5]

In his well-written book on the case, first published in 2006, Derksen argues that none of the eight 'pillars' or steps in the argument on which the conviction had been based was sufficiently solid. They did not adequately support the factual claim that Lucia had murdered or attempted to murder ten infirm patients. As Derksen stresses in his preface, he analyses the legal decision of the Court of Appeal of The Hague from the perspective of a philosopher of science, whose interest is merely to evaluate the methods and arguments used in the search for the truth. Having concluded that the evidence adduced by the prosecutors and experts did not justify a guilty verdict, Derksen also discusses psychological factors that may have induced the judges to conclude that Lucia *had* to be guilty.

For example, he analyzes four psychological tendencies that explain why we tend to commit fallacies in probability arguments: (a) the tendency to think that if there is smoke, there must be fire; (b) our inclination to think that a correlation (between the somewhat unexpected death of severely ill patients and Lucia's shifts on the ward) cannot be a coincidence; (c) the anchoring effect, by which we tend to remain focused on one scenario without sufficiently considering the many alternative scenarios that are possible, or remain focussed on a specific range of numbers; and (d) our neglect of prior probabilities.

Derksen also indicates some other circumstances that contributed to the conviction that Lucia was guilty, such as problematic aspects of her personality, and the fact that she became a nurse because she wanted to leave her earlier occupation as a prostitute. Another important factor was that a fortnight after the child Amber had died in 2001, the Juliana Children's Hospital and the Red Cross Hospital (*Rode Kruis Ziekenhuis*) issued a press release suggesting that a nurse had been responsible for the unexpected death of several patients. After the children's hospital had reported the case to the public prosecutor, and the populist daily newspaper *De Telegraaf* had published an inflamed article triggered by this press release, it seemed to be in the Dutch national interest to convict a murderess.

What were the eight evidential pillars of the factual conclusion about Lucia's guilt drawn by the Court of Appeal of the Hague in 2004? Let me summarize them briefly, using both Derksen's book and the judgment of the Court as my sources, in order to locate the Coincidence Argument in its context.

---

1   I am grateful to Colin Elliot, Richard Gill, Jeannette Leegwater, Jan-Willem Romeijn, Klaas Slooten, and to the members of the Dutch Research Seminar for Analytic Philosophy for their critical comments on, and corrections to, an earlier version of this paper. There is no need to say that all remaining errors are mine.

2   ECLI:NL:GHSGR:2004:AP2846.

3   See ECLI:NL:HR:2006:AU5496. The Supreme Court rectified the sentence of the Court of The Hague on one point only: imprisonment cannot be combined with hospital detention (*ter beschikkingstelling, tbs*).

4   ECLI:NL:GHAMS:2006:AY3864.

5   ECLI:NL:GHARN:2010:BM0876. T. Derksen, *Lucia de B. Reconstructie van een gerechtelijke dwaling*, 2006. I shall refer to the 4th edition of 2009.

First (1), the Court re-examined the cases of two seriously ill children: Amber ('victim 1'), who died at the age of six months on September 4th, 2001, and Ahmad ('victim 3'), who only weighed thirteen kilograms although he was six years old, and had a serious health crisis on January 25th of the same year. Although the death of Amber and the crisis of Ahmad had been diagnosed initially as natural occurrences by the doctors in attendance, the Court argued *in extenso* that both of the children had been poisoned intentionally by Lucia, using digoxin and chloral hydrate, respectively. As Derksen says, applying the analogy of a railway train, these two cases functioned as the 'locomotives' of the conviction, since only in these cases did there seem to be at least *some* evidence about the means Lucia de B. might have used in order to (attempt to) murder these seriously ill children, and about the times she could have applied these means.

However, having discussed in detail the medical arguments of the Court of The Hague with regard to Amber (§10.1 of the verdict), Derksen concludes that 'instead of strong, precise proof, worthy of a locomotive, we find an argument full of medical pseudo-certainties, suppressed facts, and the neglect of simple alternatives'.[6] Indeed, the girl might have died because of her many medical problems, such as her heart dysfunction, a serious brain problem, or necrotizing enterocolitis, although the court excluded this, whereas the presence of digoxin in her blood may have had a natural cause, as some experts had already argued.

Derksen's verdict with regard to the argument of the Court to the effect that Lucia had attempted to murder Ahmad (§10.3 of the verdict) is even more disconcerting. There are five quite technical contentions by the Court, each of which was necessary to justify the conclusion that Lucia had attempted to murder the seriously handicapped and mentally retarded Ahmad by giving him an overdose of chloral hydrate, a calming drug which he received anyway. Refuting only one of these contentions would have sufficed to undermine the judgment of the Court, but Derksen claims to have disproved all of them in his book. Ahmad recovered from the crisis on January 25th, and he died on February 23rd, 2001.

If both of the only two 'locomotives' of De Berk's conviction were defective, the Court of The Hague should have acquitted her of the charge of the (attempted) murders. Yet Derksen also contests the remaining seven pillars on which the Court built its verdict. Let me summarize them briefly. The second pillar (2) is the so-called Coincidence Argument, which I shall discuss more amply in this paper. According to the Court, it cannot have been coincidental that during Lucia de Berk's shifts on the ward relatively many patients had died or underwent a medical crisis of which the Court argued that natural causes could be excluded (cf. §§5.55, 11.8C, and 11.13 of the verdict). Unfortunately, however, the Court did not consider whether the number of deceased patients per year on the ward had increased when Lucia started working there. In fact, during the years 1999-2001 that she worked in the Medium Care Unit-1, six patient deaths had occurred, whereas in the years 1996-1998 before her service there were seven deaths.[7] This crucial fact already seriously undermines the Coincidence Argument, and the relevant information was available to the Court.

Step (3) in the Court's proof of Lucia's guilt is the so-called Compulsion Argument. On the day that one of the patients ('victim 7') died, Lucia wrote in her diary that she had given in to her 'compulsion' (§9.10 of the verdict). She also wrote in her diary on 28 July 1997 that she had one big secret about things she had done, and that she would take that secret to her grave (§9.2). But on 30 June 1998 she wrote that her partner had been allowed to read her diary and that she no longer had any secrets from him (cf. §9.7). Ms. de Berk consistently told the police and the courts that by the secret compulsion she meant her urge to read Tarot cards in the presence of patients in order to soothe them, which was risky because the hospital might have dismissed her for this (cf. §§9.9-13). Yet, the prosecutors and the justices interpreted the relevant passages in her diary as meaning that Lucia confessed to having a compulsion to murder (§9.20), even though many of these passages were written when no patients had died or had had a health crisis. Of course it was unfortunate that no patient could confirm De Berk's account, which was one of the grounds on which the Court rejected it (§9.18).

---

6   Derksen 2009, supra note 5, p. 77, my translation.
7   Derksen 2009, supra note 5, pp. 22, 43.

In step number (4) the Court established that five other children in the hospital had died or fallen ill unexpectedly during Lucia's shifts on the ward. Since no natural causes for these events had been traced, the Court concluded that there were none, and that there must have been an uncommon non-natural cause: murderous actions by the nurse De Berk, even though there is no indication about the nature and timing of these actions (cf. §12.5).

The Court of The Hague connected the two locomotives of step (1) to the other deaths or medical crises that occurred during Lucia's shifts on the ward by (5) the Linking Argument (*Schakel Argument*; §§5.40-5.49). Since the precise causes of these events at the relevant times were unknown, and because the evidence that Lucia had murdered or attempted to murder both Amber and Ahmad was considered to be convincing, and, finally, given the compulsion to murder allegedly testified by Lucia's diary, the Court concluded that in these other cases either a murder or an attempted murder had occurred as well. Furthermore, in steps (6) and (7) the Court argued that when three elderly and ill people at the Red Cross Hospital and the Leyenburg Hospital (*Ziekenhuis Leyenburg*) died during her shifts, Lucia must have murdered them, because the precise moment of their death had not been expected, although the doctors knew that they would die soon.

Finally (8), the Court adduced other averred evidence for Lucia's guilt, such as the facts that she contradicted herself during the interrogations, that she changed her opinion on some points, and that her written reports as a nurse were incomplete. All deaths or life-threatening incidents manifested 'a discernible and similar pattern' in that they occurred all of a sudden and unexpectedly, while in the eyes of the Court natural causes could be convincingly excluded (§§11.23-24). As Derksen shows in his 7th chapter, however, this latter conviction of the Court is highly problematic for many reasons. For example, concerning all these deaths and incidents the doctor on duty had initially given a diagnosis of a natural death or of an incident caused naturally. Even though experts disagreed about each of these cases, the Court quoted mainly those experts who supported its verdict. One cannot draw the conclusion that a natural cause did not exist from the fact that it had not been discovered. Finally, the unconditional probability that a death is due to a medical error is much greater than that it is due to a murder committed by a nurse. To what extent can hospitals be trusted to report medical errors publicly?

Although according to the psychiatric report of the Pieter Baan Centre (*Pieter Baan Centrum*), Lucia suffered from a 'complex pathological structure of her personality', consisting of a disposition to a 'strong rational control' of herself in order to conceal an underlying 'deep insecurity' and 'extreme self-hatred', the Court did not conclude that this was a case of diminished responsibility (§§14.1-2). But it may be that this report, and the facts that Lucia had attempted to commit suicide in the past, had used drugs, was a bisexual and had been a prostitute, also contributed to the Court's conviction that she must have been guilty of committing the crimes she was accused of.

I recommend to those who are able to read Dutch that they study the 18 June 2004 verdict of the Court of The Hague before reading Derksen's book. If readers did not have the background knowledge of his analysis, what would they have concluded after having read the verdict of the Court of Appeal of The Hague? Having explained its context by summarizing the verdict, I shall now focus on one argument of the Court only: the so-called Coincidence Argument.

3. The Coincidence Argument as a likelihood comparison

In its verdict of 18 June 2004, the Court of The Hague stated that it did not use any 'statistical proof' of Lucia's guilt, as the court of first instance (*Rechtbank te 's-Gravenhage*) had done (p. 1). Nevertheless, the Coincidence Argument plays an important role in supporting its conclusion that Lucia had been guilty of (attempted) murder in the eight 'non-locomotive' cases, which Derksen calls the 'wagons'. In all of these cases, a sudden and more or less unexpected death or life-threatening event occurred *during* Lucia de Berk's shifts on the ward.[8] In §11.8, the Court added that all possible natural causes of these incidents could be convincingly excluded. I shall focus first on the pure Coincidence Argument, and discuss the additional argument in Section 4.

---

8   ECLI:NL:GHSGR:2004:AP2846, §§11.8 and 11.13.

It might have been helpful if the Court of The Hague or Ton Derksen had spelled out explicitly the logical form of the Coincidence Argument, but neither of them did. In this section and the next one, I shall consider two logical reconstructions of the Coincidence Argument, and assess what can (and cannot) be shown by the argument on either of these two reconstructions. First, I construct the Coincidence Argument as a likelihood comparison (this section). However, if the argument is modelled on the Law of Likelihood, it can neither establish nor enhance a specific probability that Lucia de Berk murdered any patients.

Assuming that the Court did not use any statistical proof, such a conclusion can only be drawn if the Coincidence Argument is construed as an application of Bayes' Theorem used as a rule for updating, which will be the topic of Section 4. Constructing the argument as an application of Bayes' Theorem has the advantage of indicating all the data and premises that are needed in order to draw objectively a conclusion concerning the probability that Lucia de Berk was a murderess. Did the Court really provide these data, and did it substantiate the required premises? A logical analysis is essential for answering these questions. As I have said, I am using the case of Lucia de Berk merely as an illustration of the need to train lawyers in the logic of probability. Issues of probability play a role in nearly all arguments to the effect that a defendant is guilty.

The empirical evidence (e) on which the Coincidence Argument was based in this case is uncontested. It consists of the fact that during Lucia de Berk's shifts many more patients had died unexpectedly or had a medical crisis than during shifts by other nurses on the ward. I call this fact the fact of Concurrence, or 'Concurrence', in short. Let us now consider the following two hypotheses only (as we shall see in Section 4, many more hypotheses should be considered, and only some of these were discussed by the courts): either (h_1) this fact of Concurrence (e) was a mere coincidence, or (h_2) the Concurrence occurred because Lucia de Berk is a serial killer who murdered or attempted to murder these patients.

Clearly, evidence (e) is much more probable given the second hypothesis than given the first, at least if we assume that the lethal means a nurse would administer to her patients in order to murder them would be effective instantly during her shift on the ward.[9] This result may be expressed by the formula

$$(1) \qquad\qquad P(e|h_2) \gg P(e|h_1),$$

in which '$P(e|h_i)$' stands for the 'likelihood' of hypothesis h_i for the specific evidence e, and '$\gg$' means 'much greater than'. The 'likelihood' of a hypothesis is defined as the probability of the evidence on the assumption that the hypothesis is true.

What can we validly conclude from a premise that has the logical form $P(e|h_2) \gg P(e|h_1)$?[10] As I said, I shall first construe the argument of the Court as a likelihood argument, in which the so-called Law of Likelihood is applied. This reconstruction would be inevitable if the Court had not used any other premises in its Argument from Coincidence.

According to the Law of Likelihood, observations e *favour* hypothesis h_2 over hypothesis h_1 if and only if $P(e|h_2) > P(e|h_1)$, while the degree to which evidence e favours hypothesis h_2 over hypothesis h_1 is identical to the likelihood ratio $P(e|h_2)/P(e|h_1)$. The concept of 'favouring' used in this law is a technical notion, which involves a three-place relation between evidence e and the two hypotheses h_1 and h_2. What the Law of Likelihood says, then, is that from a premise of form (1) we can conclude that the evidence (e) favours hypothesis h_2 over hypothesis h_1 to the degree $P(e|h_2)/P(e|h_1)$, or that this evidence confirms h_2 better relatively to h_1 in the specified degree. It is important to stress that this notion of confirmation is a relative one: the evidence merely confirms one hypothesis better than the other.

---

9 Of course, a clever nurse who intends to murder a patient would use a means that would be effective later on, after her shift on the ward had ended, if at least such means were available, so that it would be more difficult to trace her crime. Although the Court of Appeal of The Hague never doubted that Lucia de Berk was a clever nurse, we do not find this consideration in its verdict (cf. Section 4, below, for further discussion). On the contrary, one of the axioms (C) of the verdict is that the charge of (attempted) homicide can only be found to have been proven if the death or medical incident had occurred at a time when the defendant was present on the ward where the patient in question was located (§5.55 of the sentence, in Dutch: '*C. het overlijden of het levensbedreigende incident moet hebben plaatsgevonden op een moment dat de verdachte op de afdeling waar de desbetreffende patiënt lag aanwezig was*'. Cf. also §11.8 of the verdict).

10 I am using the lower case letter 'h' sometimes as a constant, representing a specific hypothesis, and sometimes as a variable. Given the context, it will not be difficult to notice how it is used. If it is not possible to determine the values of $P(e|h_2)$ and of $P(e|h_1)$ unless one adds background knowledge (k), which is often the case, the formula becomes somewhat more complicated: $P(e|h_2\&k) > P(e|h_1\&k)$.

Applied to the two hypotheses we are comparing, the formula $P(e|h_2) \gg P(e|h_1)$ says that:

(2) P(Concurrence | Lucia is a serial killer) $\gg$ P(Concurrence | mere Coincidence).

In ordinary English, it states that the fact of Concurrence would be much more probable if Lucia were a serial killer than it would be on the hypothesis that it is a mere coincidence. Suppose that lemma (2) is true. What does it show? And what does it not show? It is important to stress the following points:

(a) In ordinary English, the words 'likely' and 'probable' are synonyms. But in treatises on probability theory, both terms are non-synonymous technical terms, which are defined as follows. As mentioned above, the *likelihood* of a hypothesis h with respect to evidence *e* is the probability that h confers on *e*, expressed by $P(e|h)$. In other words, the likelihood $P(e|h)$ expresses the probability that *e* occurs if hypothesis h is true, or the probability that *e* would occur if h were true. On the other hand, the 'probability' of a hypothesis h given evidence *e* is the probability that h is true if evidence *e* obtains, expressed by the formula $P(h|e)$. Given these definitions, likelihoods $P(e|h)$ are completely different from probabilities $P(h|e)$, as is also shown by the following observations.

(b) The logical properties of likelihoods as defined differ radically from the logical properties of probabilities.[11] For example, the probabilities of a hypothesis (h) and of its negation (not-h) always add up to 1, that is, it is certain that either h or not-h is true, whatever the evidence: $P(h|e) + P(\text{not-h}|e) = 1$ for all items of evidence *e*. But the likelihoods $P(e|h)$ and $P(e|\text{not-h})$ may not add up to 1. For example, it is highly unlikely that you become president of the United States (*e*), both if (h) you are an American and if (not-h) you are not an American, as someone might have said to Barack Obama in 2004, so that in this case $P(e|h) + P(e|\text{not-h}) \ll 1$.[12]

Furthermore, a logically stronger (richer in content) hypothesis is less probable than a logically weaker hypothesis, but they may have the same likelihood with regard to a specific piece of evidence *e*. For example, the hypothesis (h_1) that at the next round of card dealing you will get a ten of spades is stronger than hypothesis (h_2) that you will get a spades card ranked higher than the six of spades, because h_1 entails h_2 and not *vice versa*, so that it is less probable that (h_1) will turn out to be true than that (h_2) will turn out to be true. Suppose that your next card is dealt, and you see that the wristwatch of the dealer reflects a spades symbol, which is your evidence *e*. Clearly, the likelihoods $P(e|h_1)$ and $P(e|h_2)$ are the same in this case, because if this is the evidence (*e*), $P(e|h_1) = P(e|h_2) = (\text{nearly})1$.

(c) Since probabilities $P(h|e)$ and likelihoods $P(e|h)$ are logically and conceptually very different, it is fallacious to conclude from a comparison of likelihoods $P(e|h_2) > P(e|h_1)$ that the same holds for the probabilities $P(h|e)$, so that it would follow that $P(h_2|e) > P(h_1|e)$. Applied to our example, it would be fallacious to conclude from the likelihood comparison (2) P(Concurrence | Lucia is a serial killer) $\gg$ P(Concurrence | mere Coincidence) that it is more probable that the Concurrence was caused by a murderous nurse than that it happened accidentally. The inference that if the truth of a hypothesis would make a specific fact extremely improbable, this fact, if established, makes it improbable to the same extent that the hypothesis is true, has been called the Prosecutor's Fallacy, because it is committed by prosecutors so often. The following logical analogy will illustrate further that such an argument is fallacious.

Suppose, for example, that (*e,* imagined) a well-known American atheist caught Ebola on October 20, 2014, when travelling on subway line E in Manhattan. As far as we know, there was only one other Ebola patient in New York at that time, the doctor Craig Spencer, who had worked with Doctors Without Borders in Guinea, and who had travelled on several subway lines before he fell ill and was tested positive for the deadly disease. Clearly, the likelihood $P(e|h_1)$ that the atheist was infected (this is evidence *e*) accidentally by Dr. Spencer, because both of them travelled by subway in Manhattan, which is hypothesis (h_1), is extremely small for many reasons. Formulated in terms of the formula: P(the atheist was infected | Craig Spencer and the atheist travelled by subway in Manhattan) is near to 0.

---

11 By definition, all probabilities and likelihoods lie between 0 (impossible) and 1 (certain). So, the probability of a necessary truth equals 1, such as 'either *p* or not-*p*', and the probability of a necessary falsehood equals 0.

12 In this sentence, I am not using the term 'unlikely' in the technical sense of the term, as terminological purists will have observed. Strictly speaking, the term should be applied only to the hypothesis h in the formula $P(e|h)$, and not to the evidence *e*. However, in order to avoid confusion I shall sometimes apply the term 'likely' to *e* when I am contrasting likelihoods $P(e|h)$ with probabilities $P(h|e)$.

When the *New York Times* reported the case, however, a Christian Radical minister instantly launched another hypothesis (h_2) in order to explain fact (*e*) that the well-known atheist caught the Ebola virus: God wanted to punish him for his atheism, and, being omnipotent and able to cause a miracle, God did so by creating the virus out of nothing in the atheist while he was travelling on subway line E. Clearly, the likelihood comparison $P(e|h_2) \gg P(e|h_1)$ holds, because $P(e|h_2) = 1$ whereas $P(e|h_1)$ is near to zero. Yet no reader will conclude, I hope, that given the evidence (*e*), the second hypothesis (h_2) is more probably true than the first. Even though the evidence strongly favours the second hypothesis over the first, from the logical point of view *this does not imply anything* concerning the probabilities that one of these hypotheses is true. The same holds true for the Coincidence Argument concerning Lucia de Berk, if it is construed as a likelihood comparison.

(d) In order to see this even more clearly, it may help to remember that extremely small likelihoods often occur, so that one cannot infer validly a small probability $P(h|e)$ from a small likelihood $P(e|h)$. For example, the likelihood that one gets a specified particular deck of cards when playing bridge given the chance hypothesis that the cards were dealt fairly (h_1) is approximately $1/5.36 \times 10^{28}$, that is: 1 divided by $52!/(13!)^4$, or 1/53,644,737,765,488,792,839,237,440,000, since there are 52 cards and 13 rounds of dealing. The likelihood of the alternative hypothesis that you got your deck because an omnipotent evil demon manipulated the dealing process without being noticed, and intended to give you precisely this quite bad deck (h_2), is much greater, to wit: (nearly) 1. So in this case, $P(e|h_2) \gg P(e|h_1)$, but again, nobody will conclude concerning such an obvious example that it is more probable that the evil demon hypothesis is true than that the chance hypothesis is true.

(e) At this point, I should remind the reader once again that the Court of Appeal of The Hague decided *not* to use any statistical calculations, as the court of first instance had done when it relied on the expertise of Professor Henk Elffers. According to the calculations by Elffers, the chance that Lucia during her 142 shifts at the Juliana Children's Hospital in The Hague (out of the 1029 shifts by nurses that took place during the period she worked at the hospital) had been present by accident at eight medical incidents, which were all the medical incidents that occurred during this period according to the data provided by the hospital, was about 1 in 9,000,000. Multiplying this fraction with the fractions obtained by similar calculations for the other two hospitals where Lucia had worked, Elffers estimated the chance that Lucia had been present purely by accident at the medical incidents that occurred during her shifts at 1 in 342,000,000. The null hypothesis that Lucia's presence at these medical incidents was a pure coincidence should be rejected if one uses as a test of significance a probability threshold of 1/10,000, as Elffers decided to do. Of course, rejecting the null hypothesis does not entail that Lucia was a murderess, because there are many other possible explanations for her presence during relatively many medical incidents, as Elffers stressed, and as we shall see in the next section.

It was wise of the Court of The Hague not to use such statistical calculations, since many experts contested the calculations by Elffers. For example, in his book on the case Ton Derksen rejects Elffers' assumption that the chances obtained for each of the three hospitals should be multiplied with each other, because applying this rule for calculating their sums would imply that the chance that a nurse is present accidentally during the same number of incidents would be much lower if she changed hospital than if she did not change hospital.[13] Using a number of other arguments, Derksen concludes that the chance that a nurse would work on the ward coincidentally during the number of incidents that Lucia experienced might be 1/44, and he claims that the statistician Richard Gill concluded an even higher chance: 1/9.[14] If these latter calculations are only moderately plausible, it follows that one cannot use any statistics as an argument for rejecting the null hypothesis of coincidence. For they imply that a similar coincidence of medical incidences and the shifts of one nurse might occur every year at some Dutch

---

13 Derksen 2009, supra note 5, pp. 140-142.
14 Derksen 2009, supra note 5, p. 144. Unfortunately, in note 48 to his page 144, no publication by Richard Gill is mentioned. Gill's calculations were based upon data that Derksen provided, from which 'incidents' were removed for which De Berk was not judged guilty. But from a logical perspective, the statistics should be based upon an objective and medical definition of 'incident', which is independent of conclusions about Lucia's guilt. When Gill discovered the biased nature of Derksen's data, he calculated a chance of 1 in 26. See: R.D. Gill et al., 'Elementary Statistics on Trial (the case of Lucia de Berk)', 2010, available at: <http://arxiv.org/abs/1009.0802> (last visited 16 January 2015). I am grateful to Professor Gill for this information.

hospital. In other words, the results do not pass the test of significance, which, as Elffers stressed, should be quite demanding in this context.

We must conclude, then, that the argument to the effect that Lucia's presence during relatively many medical incidents *cannot* have been a coincidence, is invalid, not only if it is construed as a likelihood comparison, but also if it had been based upon statistical calculations.

4. A Bayesian version of the Coincidence Argument

As we have seen in Section 3, the Argument from Coincidence is a fallacy if its logic is construed as an application of the Law of Likelihood. One might conclude from this result that we should try to reconstruct the argument differently, if possible, because it would be uncharitable to endorse a logical analysis of the argument on which it is invalid. One requirement for a more charitable reconstruction is that the Coincidence Argument would really enhance the probability that the murderess hypothesis is true. Another requirement is that we might validly conclude to a specific value of this probability, whether specified quantitatively or not, on the basis of all evidence for and against. For, surely, a criminal court is allowed to convict the accused of a series of murders only if the hypothesis that she is a murderess is very probably true given all the relevant evidence (for and against!), that is, if $P(\text{murderess} \mid \text{evidence}_{1\text{-}n}) \gg 1/2$.

As many experts have argued concerning the case of Lucia de Berk, the best way to formalize the Argument from Coincidence in order to meet these requirements is by using the Rule or Theorem of Bayes, called after Thomas Bayes (1702-1761), a non-conformist English Presbyterian minister and statistician, who first derived it. Bayes' Theorem may be formulated as follows:

$$(3) \qquad\qquad P(\text{h}|e) = \frac{P(e|\text{h}).P(\text{h})}{P(e)} \,.\text{[15]}$$

It says that the probability P that a hypothesis h is true given new evidence e is equal to the likelihood of h with regard to e multiplied by the prior probability that the hypothesis is true, and divided by the probability that the evidence obtains whether or not the hypothesis is true. Clearly, if and only if $P(e|\text{h}) > P(e)$, the evidence e *enhances* the original or 'prior' probability $P(\text{h})$, so that only in this case, $P(\text{h}|e)$ will be greater than $P(\text{h})$. With regard to evidence e, $P(\text{h}|e)$ is called the posterior probability of the hypothesis, and $P(\text{h})$ is called the prior probability, that is, the probability that the hypothesis is true before evidence e is taken into account.

If one wants to derive a specific probability $P(\text{h}|e)$ that the hypothesis is true, one should start with a specific prior probability $P(\text{h})$, except in cases in which the evidence accumulates so strongly that it 'swamps' possible differences in the value of the prior probability $P(\text{h})$.[16] However, in the case of Lucia de Berk, this does not obtain, because the evidence was not that overwhelming, to say the least. It follows that in this case, one cannot derive a value of the probability that Lucia was a murderess (whether quantified precisely or not), unless one provides a value for the prior probability $P(\text{h})$. I shall discuss below (under (b)) the issue of how this might be done. But neither the court of first instance, nor the Court of Appeal of The Hague, even attempted to do so in their decisions.

We should distinguish between Bayes' Theorem, on the one hand, and Bayesianism as a normative theory of epistemic rationality, on the other. As a theorem of probability theory, Bayes' Theorem is uncontroversial. It can be easily derived from the axioms of probability theory, which are the basic rules of consistency for assignments of probability, if one adds Kolmogorov's definition of conditional probability $P(\text{h}|e) = P(\text{h}\&e)/P(e)$.[17] According to Bayesianism as a theory of epistemic rationality, rational subjects have degrees of belief. These degrees of belief should be interpreted as probabilities, and they should be updated by applying Bayes' Theorem or Rule.

---

15 Again, I should stress that often we need background knowledge k in order to determine the value of terms such as $P(e|\text{h})$. If so, we might spell out Bayes' Theorem as $P(e|e\&k) = P(e|\text{h}\&k).P(\text{h}|k)/P(e|k)$.

16 Cf. W. Salmon, *Reality and Rationality*, 2005, p. 81.

17 Cf., for example, E. Sober, *Evidence and Evolution. The Logic behind the Science*, 2008, p. 9, or C. Howson & P. Urbach, *Scientific Reasoning. The Bayesian Approach*, 2nd edition, 1993, pp. 27-29.

Suppose that we are able to establish evidence *e* beyond doubt, so that we do not need to apply any probability estimates to the statement that evidence *e* obtains. Then we can apply what is called 'strict conditionalization': as soon as we have established that *e* obtains, we update our original estimate of the prior probability that hypothesis h is true by applying Bayes' Theorem, that is, we calculate the posterior probability $P(\text{h}|e)$ that the hypothesis is true. In criminal cases strict conditionalization is often inappropriate, because with regard to many pieces of evidence it cannot be established beyond any doubt that they obtain. If so, one has to apply probability estimates to statements about the evidence as well, so that the rule for updating will be more complicated. In the following discussion of the Argument from Coincidence, however, I shall avoid these complications. Let us assume for the sake of simplicity that the fact of Concurrence has been established beyond doubt.[18]

If we formulate the Coincidence Argument in terms of Bayes' Rule, the argument takes the following form:

(4) $\quad P(\text{Lucia is a serial killer} \mid \text{Concurrence}) =$

$$\frac{P(\text{Concurrence} \mid \text{Lucia is a serial killer}) \cdot P(\text{Lucia is a serial killer})}{P(\text{Concurrence})},$$

in which the fact of Concurrence updates the prior probability that we attached to the serial-killer hypothesis. As I said, the advantage of construing the Coincidence Argument as an application of Bayes' Rule is that this specifies clearly all data and supporting arguments required in order to draw a conclusion about the probability that Lucia de Berk really was a murderess. Let me now spell out what is needed in order to draw such a conclusion by focusing in turn on each of the terms of the formula at its right-hand side.

(a) The Court of The Hague would not have used the Coincidence Argument in its verdict if it had not assumed that the Concurrence (that is: the fact that during Lucia's service on the ward relatively many patients had died or had a health crisis) would be quite likely if the murderess hypothesis were true. As we saw, evidence *e* only enhances the prior probability of the hypothesis if $P(e|\text{h}) > P(e)$. Let me start by focusing on the first term of this inequality, to wit $P(e|\text{h})$. How should we determine the likelihood $P(\text{Concurrence} \mid \text{Lucia is a serial killer})$?

As I said above in a footnote, it is not at all self-evident that this likelihood is near to certainty, which the courts assumed. I would rather estimate the likelihood $P(\text{Concurrence} \mid \text{Lucia is a serial killer})$ as not above 1/10, for the following reasons. One piece of background knowledge mentioned by the Court is that nurse Lucia was not stupid.[19] The Court of Appeal even averred: 'in her actions, the suspect proceeded in an extremely sophisticated and methodical manner, so that the risk was small that her crimes would be discovered'.[20] Suppose, then, that she was both clever and a serial murderess. If she had premeditated her acts, as is necessary for the verdict of 'murder', would she have used methods of murdering that are operative instantaneously? I do not think so, because in that case she would have foreseen that the patients would die during her shift on the ward, so that she risked becoming a suspect. In other words, the fact of Concurrence might just as well be used as an argument *against* the murderess hypothesis, if at least in principle lethal means were available that do not cause death instantaneously. Instead of enhancing the prior probability of the hypothesis, the fact of Concurrence might rather diminish it. This is a first reason why, even on the Bayesian reconstruction, the Argument from Coincidence is unconvincing, although it plays a crucial role in the decisions of the courts.[21] Surprisingly, neither the court of first

---

18 Many types of doubt may be raised as to whether the fact of Concurrence has been established really and with precision. For example, the courts did not make an inventory of all the medical incidents that occurred when Lucia was not present on the ward. Cf. Derksen 2009, supra note 5, p. 122.

19 According to the report of the Pieter Baan Centre of 28 February 2003, Lucia de Berk is an 'intellectually gifted woman' ('*een intellectueel begaafde vrouw*'): §14.2 of the verdict of the Court of Appeal of The Hague.

20 Verdict of the Court, §15.5: '*De verdachte is bij haar handelen uiterst geraffineerd en planmatig te werk gegaan waardoor de kans op ontdekking van haar misdaden gering was*' (my translation in the main text). The Court made this claim in order to explain why it was impossible with regard to most deaths and medical incidents to find *any* evidence that supported the serial-killer hypothesis, such as evidence about the means used. But it overlooked that the same claim might undermine its Coincidence Argument.

21 To my amazement, I did not find this first reason in the literature on the case of Lucia de Berk, not even in Derksen's book. Let me remind

instance nor the Court of Appeal of The Hague reflected critically on their assumption that the likelihood P(Concurrence | Lucia is a serial killer) is very high, at least higher than $P(e)$, the second term of our inequality. Let me discuss this second term later (under (c)), because it will turn out to conceal many complexities. I focus first on the prior probability of the murderess hypothesis.

(b) What is the prior probability P(h) that a nurse such as Lucia de Berk is a serial killer, and how should we establish this probability? As I said, one needs to assume a specific prior probability in order to derive the posterior probability P(h|e) on the basis of the empirical evidence e, because in this case the prior is not 'swamped' by the evidence. A crucial lesson of Bayesianism may be expressed by the maxim: 'no probabilities out without some probabilities in'.[22] But how should we determine the prior probabilities we plug into applications of Bayes' Theorem? There are two options here, which are both problematic to some extent, and which have been discussed extensively in the literature.

First, one might hold the 'subjectivist' view that the prior probability P(h) merely reflects our subjective estimate of the probability that h is true.[23] According to this subjectivist interpretation, the rule of Bayes is used merely as a logically conclusive method for updating, on the basis of new evidence, the strength of our subjective conviction concerning a hypothesis or a belief. However, if probability arguments are used by criminal courts in order to establish beyond reasonable doubt that a defendant is guilty, one cannot use this purely subjectivist interpretation of Bayesian updating. The conclusion of the probability argument that on the basis of all evidence e the probability P(h|e) is sufficiently high for conviction in a criminal lawsuit, can only be drawn if the value of the prior probability P(h) has been established more objectively, on the basis of good arguments and evidence.

So let us ask once again: what is the prior probability that a nurse such as Lucia de Berk is a serial killer? Experts have reproached the Court of The Hague in that it did not establish the prior probability of the murderess hypothesis explicitly and somewhat objectively. This is a serious issue, as I said, because the empirical evidence used for updating the prior probability that Lucia was a murderess, such as the fact of Concurrence, is not very convincing. The less convincing the evidence is, the higher should be the prior probability in order to conclude that the defendant is guilty beyond reasonable doubt. And the lower the prior is, the more compelling empirical evidence one would need concerning the means used to murder, the motives for doing so, and the times and opportunities for applying these means. However, such compelling evidence was clearly lacking in all the 'non-locomotive' accusations of Lucia de Berk.

In order to establish the prior probability of the murderess hypothesis somewhat objectively, we should use frequency data, at the least. Psychologists such as Kahneman and Tversky have discovered that we often neglect frequencies in estimating prior probabilities, at least if we are focused on the specific salient properties of an individual. For example, during an opening ceremony of the academic year at a university I am visiting for the first time, I see a weird professor with a long beard and piercing eyes, who is not known to me. A colleague asks: is he a professor of philosophy or of law? If I answer on the basis of his looks: 'He must be a philosopher', I commit what is called a Base Rate Fallacy. At this university, sixty of the three hundred professors are professors of law, whereas there are only two professors of philosophy. Hence, the prior probability that the unknown professor is a philosopher is only 1/150, whereas the prior probability of him being a professor of law is much larger: 1/5. In this example, I overestimate a posterior probability because I neglect the prior probability on the basis of salient characteristics. Hence I also commit a Fallacy of Salience.

In the case of Lucia de Berk, the courts should have investigated explicitly the frequency that a nurse such as Lucia is a serial killer, in order to avoid the risk of a Base Rate Fallacy or a Fallacy of Salience.[24] The risk of such fallacies is considerable in this case, because of Lucia's past as a prostitute, for example. Suppose we choose as a relevant population the number of nurses (male and female) in the Netherlands. What is the frequency of serial murderers among them? Let us assume that in the year 2000 there were

the reader of axiom (C) of the verdict of the Court of Appeal of The Hague: that the charge of (attempted) homicide can be found proven *only if* the death or medical incident occurred at a time at which the defendant was present on the ward where the patient in question was located (§5.55, my emphasis, cf. §11.8).

22   Sober 2008, supra note 17, p. 21.
23   Subjectivism in this sense should not be confused with subjective interpretations of the notion of probability itself. In this latter sense, probabilities are degrees of belief, as Bayesianism assumes.
24   Cf. Derksen 2009, supra note 5, pp. 38-44 for some exemplary calculations of the prior.

200,000 nurses in this country (I could not find reliable statistics on this matter). Let us also assume that we can establish on the basis of reliable data that among Dutch nurses there has been one serial murderer during the last fifty years. If, on average, serial murderers are caught within two years, the chance that a serial murderer is active in an arbitrarily selected year is 1/25. Assuming that the number of nurses has been constant, we might estimate the prior probability that a specific nurse is a serial killer at 1/(25 x 200,000) = 1/5,000,000. If we apply this frequency to the Lucia case, the prior probability $P(h)$ of the serial-killer hypothesis is 1/5,000,000. Clearly, then, quite a lot of convincing evidence would be needed in order to conclude legitimately that Lucia was a serial killer!!!

The numerator of Bayes' Theorem consists of the likelihood times the prior probability of hypothesis h: $P(e|h).P(h)$. Let us now calculate the value of this numerator if applied to the Argument of Coincidence in the case of Lucia de Berk, using the assumed numbers by way of illustration. I estimated the likelihood P(Concurrence | Lucia is a serial killer) roughly at 1/10 on the basis of the argument under (a). Hence, the resulting probability of the numerator P(Concurrence | Lucia is a serial killer).P(Lucia is a serial killer) will be around 1/50,000,000. If the Argument from Coincidence were the only argument used, it would follow that the probability of the denominator P(Concurrence) should be less than one in 25 million if the resulting probability $P(h|e)$ that Lucia is a serial killer should be above one half, which is a minimalistic requirement for a conviction.

One might think that this condition is satisfied if statistician Elffers is (mis-)interpreted as estimating the likelihood P(Concurrence | mere Coincidence) at one in 342 million. But as we shall see under (c), this would be yet another logical blunder. The reason is that the coincidence hypothesis is not the only hypothesis we should consider in order to determine the value of the denominator P(Concurrence). In other words, we cannot equate this denominator with the likelihood P(Concurrence | mere Coincidence). Furthermore, one should take into account the totality of the evidence for and against available to the Court in order to calculate the resulting probability $P(h|e)$. Before discussing the denominator of Bayes' Theorem in this case, however, an objection may be raised.

As Derksen also does in his book, we assumed that the relevant population for determining the frequency that an individual like Lucia de Berk is a serial killer is the population of nurses in the Netherlands.[25] But why should we opt for this population as a reference class? In its verdict, the Court of The Hague mentions some expert information about Lucia de Berk's psychology. The structure of her personality was 'complex and pathological', because she concealed her deep uncertainty and self-hatred by 'rigid rational self-control'.[26] If so, should we not attempt to determine the frequency of female serial killers in a population of persons with a personality structure and a problematic childhood development such as Lucia's? One might also apply Hickey's Trauma Control Model, which explains how an early childhood trauma might set up the child for deviant behaviour later on, since, allegedly, Lucia suffered from such a trauma. It is questionable, however, whether this choice of a reference class would result in a higher frequency. Statistics show that female serial killers are very rare in comparison to their male counterparts, whereas they typically kill their husbands or lovers. According to one estimate, there were only 64 known female serial killers in the U.S. between 1800 and 2014.[27]

My point here is purely an epistemological one: it may be somewhat arbitrary how one chooses the reference class for determining the frequency of serial killers. Arguably, the narrowest population should be considered from which Lucia may be regarded as being drawn at random. However, this is only one of the problems for determining the prior probability that Lucia de Berk was a serial killer. Another problem is that frequencies do not imply anything about a particular instance, since they merely quantify the relation of a subclass of instances to the reference class. It will not be easy, then, to establish the prior probability that Lucia de Berk was a serial killer on the basis of compelling empirical evidence. Without any attempt to do so, however, a conviction cannot be convincing, as Bayes' Theorem indicates, unless the evidence swamps the prior.

---

25   Derksen 2009, supra note 5, pp. 40-42.
26   Verdict of the Court, §14.2.
27   Wikipedia article on 'Serial Killer', consulted on 11 November 2014, and A. Frei, et al., 'Female serial killing: Review and case report', 2006 *Criminal Behavior and Mental Health* 16, no. 3, pp. 167-176.

(c) Let us now consider, finally, the third term at the right-hand side of Bayes' Theorem: the denominator $P(e)$. As we have seen, the evidence of Concurrence enhances the prior probability that Lucia is a serial killer only if $P(e|h)$ is larger than $P(e)$. How should the value of the latter term be established? This may seem to be a simple question, because at first sight the $P(e)$ is a simple term of the equation. However, this impression is misleading, since the term conceals several complexities.

In ordinary English, we might formulate what $P(e)$ means as follows: the probability of the evidence whether hypothesis h is true or not. Since if h is not true, not-h (written as ~h) must be true, $P(e)$ may be spelled out as follows:

$$(5) \qquad\qquad P(e) = P(e|h).P(h) + P(e|\sim h).P(\sim h).$$

We have already discussed the first part of the right-hand side of this identity, to wit $P(e|h).P(h)$, which is the numerator in Bayes' Theorem. But what, exactly, does the symbol '~h' stand for? How should we determine the values of the likelihood $P(e|\sim h)$ and of the prior probability $P(\sim h)$? Applied to the Coincidence Argument as construed in Bayesian terms, '~h' stands for the hypothesis that the Concurrence is *not* due to Lucia being a serial killer. But if the Concurrence is not due to such a horrendous cause, it must be due to something else. Many alternative hypotheses might be put forward at this point, so that the negation of h is often called a 'catch-all hypothesis'. Spelled out in a formula, this means that:

$$(6) \quad P(e) = P(e|h_m).P(h_m) + P(e|h_c).P(h_c) + P(e|h_1).P(h_1) + P(e|h_2).P(h_2) + P(e|h_3).P(h_3) + \ldots\ldots P(e|h_n).P(h_n),$$

where h_m stands for the murderess hypothesis, h_c stands for the hypothesis of mere coincidence, and 'h_1', 'h_2', 'h_3', etc. stand for rival explanations of the fact of Concurrence.

We now see, as promised, that the simple formula $P(e)$ conceals an impressive complexity. In order to determine its value, we should add up the likelihood times the prior probability of h_m and of all these alternative hypotheses. Let us formulate some of them, in order to stimulate the reader's imagination.

As expert Elffers already indicated, the fact e of Concurrence may be explained by supposing that:

(h_c) the Concurrence was a mere coincidence;

(h_1) at the times of Concurrence, Lucia always shared her shifts with someone else, who caused the incidents;

(h_2) Lucia was often on a night shift, and the risk of incidents is higher during the night;

(h_3) Lucia is a relatively incompetent nurse, so that the risk of incidents during her shift on the ward is high;

(h_4) Lucia prefers to care for patients with complex disorders, and these patients have a greater risk of dying;

(h_5) Lucia prefers to care for patients who are seriously ill;

(h_6) someone hates Lucia and tries to discredit her.[28]

In his discussion of the Argument from Coincidence, Derksen adds two other hypothetical explanations of the fact of Concurrence:

(h_7) patients felt more at ease during Lucia's presence, and they die more easily when they are relaxed;

(h_8) if a patient dies or has a crisis during Lucia's shift on the ward, this receives more attention than similar incidents during shifts by other nurses, and it will be classified or reclassified more easily as non-natural.[29]

We might invent further hypotheses in order to explain the evidence of Concurrence, such as:

(h_9): as is clear from the testimonies of other nurses, Lucia's personality resulted in strong feelings of sympathy or antipathy among her colleagues. Suppose that this is also true with regard to her patients, and suppose that the arousal of strong feelings in patients of certain types may evoke a medical crisis. Or:

---

28 H. Elffers, *Rapport Verdeling reanimatie- en overlijdensgevallen in het Juliana Kinderziekenhuis en het Rode Kruisziekenhuis*, May 8th, 2002, p. 7, quoted by Derksen 2009, supra note 5, p. 125.
29 Derksen 2009, supra note 5, pp. 125-126.

(h_{10}): in the critical cases of Amber and Ahmed, medical errors played a crucial role. It was tempting for the medics involved to distract attention from medical errors by supporting the accusation of Lucia de Berk, who was not popular among her colleagues because of her past and personality. They did so by constructing the evidence of Concurrence artificially. Yet another hypothesis might be:

(h_{11}): at least one of the causal factors mentioned in hypotheses h_{1-10} was operative in each of the instances of Concurrence.

Both Elffers and Derksen stress that the rejection of the hypothesis of coincidence (h_c) does not imply that the serial-killer hypothesis is true, because there are so many other possible explanations for the fact of Concurrence. The Court of Appeal thought that it could exclude hypothesis h_4, or rather h_5 (§§11.14-16), and hypothesis h_2 (§§11.17-18). It also rejected hypothesis h_3 that Lucia was not a competent nurse, relying mainly on Lucia's testimony that she was not incompetent (§§11.19-21), and it had already shown that hypothesis h_1 was false. But of course, as Derksen observes, it does not follow that the serial-killer hypothesis is true or even confirmed, because not all alternative explanations of the fact of Concurrence had been refuted.[30] In particular, one might suppose that both the likelihood and the prior probability of h_7 are quite high if Lucia was often laying Tarot cards, as she testified with regard to the 'compulsion' passages in her diary. As she said, by laying Tarot cards in their presence she aimed to sooth her patients. And, of course, a hypothesis such as h_{10} should have been investigated thoroughly by the Court of The Hague, since medical errors are much more common than nurses who are serial killers.

However this may be, introducing Bayes' Theorem explicitly, and spelling out the prior probability of the evidence as done in lemma (6), shows that the *onus probandi* with regard to the Argument from Coincidence is much more subtle than the Court of The Hague presumed. One cannot simply reject all alternative explanations of the fact of Concurrence, and conclude that Lucia must be a serial killer. What one should do is weigh carefully the prior and the likelihood of each of the competing hypotheses, then multiply them with each other, and add up the results, in order to calculate the value of $P(e)$. Of course, it will not be possible to obtain precise numerical values for each of these factors, so that one has to proceed intuitively. Let me illustrate this by discussing what the Court of The Hague said about hypothesis h_4, according to which Lucia preferred to care for patients with complex disorders, and about hypothesis h_5, which says that she preferred to care for seriously ill children.

In its §§11.15 and 11.16, the Court merely quoted Lucia's pronouncements. Interestingly, some of the passages cited by the Court seem to confirm hypothesis h_4, whereas others disconfirm it. For example, Lucia stressed during the court session of March 18th, 2004, that she tried to avoid children who were seriously ill, even though this was not always possible. This statement disconfirms hypothesis h_5, which the Court rejected in §11.14 of the verdict. She then added: 'I preferred children who needed more complex care.'[31] But during the court session of March 22nd, 2004, she said that although she often had to care for 'complex children', this was also true for her colleagues.[32]

How should we evaluate $P(e|h_4).P(h_4)$ on the basis of these considerations? It seems that Lucia made a distinction between seriously ill children on the one hand, and complex cases on the other. She tried to avoid the former and preferred the latter. Suppose, however, that complex cases also run a greater risk of dying or having a crisis, as hypothesis h_4 suggests. One might conclude that the fact of Concurrence would have been quite likely if hypothesis h_4 were true, so that one might evaluate the likelihood $P(e|h_4)$ in the order of magnitude of 1/10. Suppose that one evaluates the prior $P(h_4)$ not as high as Lucia's statement of March 18th suggests, because of her statement of March 22nd, so that we value it in the order of magnitude of 1/5. It follows that we should value $P(e|h_4).P(h_4)$ in the order of magnitude of 1/50.

It is important to stress that this is only one of the many resulting numbers that have to be added up in order to get the value of $P(e)$. Clearly, if by adding up all these numbers we would arrive at 1/10, which is not implausible, the Coincidence Argument would not even enhance the prior probability that Lucia is a serial killer, since we estimated the likelihood P(Concurrence | Lucia is a serial killer) at 1/10 on the basis of the argument under (a). As explained above, evidence e enhances the prior probability $P(h)$ only if $P(e|h) > P(e)$. In other words, if this would be the value of $P(e)$, the Argument from Coincidence would

---

30 Derksen 2009, supra note 5, p. 127.
31 Verdict of the Court of The Hague, §11.15: '*De kinderen die een complexere zorg behoefden hadden mijn voorkeur.*'
32 Verdict of the Court of The Hague, §11.16.

be simply irrelevant. How can we square this possibility with the fact that the Coincidence Argument played such a crucial role in the verdict of the Court of Appeal? The most plausible explanation of this incongruity is, without any doubt, that the justices were not trained sufficiently, or not at all, in the logic of probability.

5. Conclusion

In this article, I discussed two possible logical reconstructions of the Coincidence Argument, which played an important role in the conviction of Lucia de Berk in 2004. If the argument is constructed as an instance of the Law of Likelihood, nothing follows concerning the probability that Lucia was a murderess. If, however, the Argument of Coincidence should be seen as an instance of Bayesian updating, the Court of The Hague did not fathom the diversity of data needed in order to make it work.

Of course, the Argument from Coincidence was not the only argument the Court of The Hague adduced. I leave it to the reader to apply a Bayesian analysis to the other arguments, such as the Linking Argument (*Schakel Argument*) or the Compulsion Argument. In order to draw a justified conclusion concerning the probability that Lucia was a serial killer, all evidence for and against should be considered, including the many pieces of evidence against which were neglected by the Court. In other words, one should apply what philosophers call the 'principle of total evidence'. I mentioned one important piece of evidence that the Court neglected: the simple fact that when Lucia worked in the Medium Care Unit-1, the number of unexpected deaths did not increase at all. That would indeed be surprising if she really was a clever serial killer! In other words, the likelihood of the murderess hypothesis with respect to this fact is indeed low.

As has been argued by several statisticians and also by Ton Derksen, one cannot draw a conclusion about the probability that Lucia was guilty from the evidence adduced, unless one also specifies the prior probability of the murderess hypothesis. This should be done with some objectivity. The purely subjectivist interpretation of Bayesian updating has to be rejected if Bayesianism is used in criminal procedures. I indicated two problems concerning the use of frequencies for determining prior probabilities.

The main conclusion of this paper is that courses on legal reasoning should contain an introduction to the logic of probability, since probabilities often play a role in arguments concerning matters of fact. The well-known case of Lucia de Berk is a salient illustration of this claim. Let me finish with a quote from *Practitioner Guide No. 1* by Colin Aitken et al. Since '[s]tatistical evidence and probabilistic reasoning (…) play an important and expanding role in criminal investigations, prosecutions and trials (…), [i]t is vital that everybody involved in criminal adjudication is able to comprehend and deal with probability and statistics appropriately. There is a long history and ample recent experience of misunderstandings relating to statistical information and probabilities which have contributed towards serious miscarriages of justice.'[33] ¶

---

33 C. Aitken et al., *Communicating and Interpreting Statistical Evidence in the Administration of Criminal Justice. Practitioner Guide No 1. Fundamentals of Probability and Statistical Evidence in Criminal Proceedings. Guidance for Judges, Lawyers, Forensic Scientists and Expert Witnesses*, 2010 (<www.rss.org.uk/Images/PDF/influencing-change/rss-fundamentals-probability-statistical-evidence.pdf>, last visited 19 January 2015), p. 3.