

University of Groningen

## Risks of Incorrect Use of Probabilities in Court and What to Do about Them

Mackor, Anne Ruth

*Published in:*  
Risk and Responsibility in Context

*DOI:*  
[10.4324/9781003276029-7](https://doi.org/10.4324/9781003276029-7)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2024

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Mackor, A. R. (2024). Risks of Incorrect Use of Probabilities in Court and What to Do about Them. In A. Placani, & S. Broadhead (Eds.), *Risk and Responsibility in Context* (pp. 94-108). (Routledge Studies in Ethics and Moral Theory). Routledge. <https://doi.org/10.4324/9781003276029-7>

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

*Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.*

# 5 Risks of Incorrect Use of Probabilities in Court and What to Do about Them

*Anne Ruth Mackor*

## 5.1 Introduction

Over the past decennia, forensic evidence, such as DNA, fingerprints, and gunshot residue, has come to play an increasingly important role in criminal cases. As a consequence, nowadays at least three different kinds of expertise are called upon in judicial evidential decision-making. First, expertise is required with respect to the law. The selection and interpretation of the relevant legal rules and the selection and qualification of legal facts are the “core business” of judges who are trained in law. Alongside that, expertise is required with respect to the proof of those facts that judges seek to qualify as legal facts. For example, in order to decide whether a defendant committed manslaughter, the court must determine, among others, whether stabbing was the cause of death of the victim and whether it was the defendant who stabbed the victim. In doing so, courts increasingly rely on the expertise of the forensic sciences. Forensic scientists present their findings about the evidence in terms of degrees of probability, more specifically in terms of likelihood ratios. Therefore, a third type of expertise has become increasingly important in legal cases, namely expertise in statistics and Bayesian probability theory.

In most Western countries, the judiciary consists solely of jurists. Forensic scientists, statisticians, and probability theorists are not members of the court. Instead, courts can call upon these expert witnesses as advisors.<sup>1</sup> The introduction of these expert witnesses in courts is meant to improve judicial evidential decision-making, but it also introduces the risk of judicial misunderstandings and misapplications of forensic findings. Accordingly, we are confronted with the paradoxical situation that the introduction of expert witnesses in court can cause the quality of the court’s evidential decisions to deteriorate instead of improve.

The case of *R v. Sally Clark* is perhaps the most infamous example of the adverse effects of forensic experts using statistics and probability theory in court.<sup>2</sup> A medical expert witness made statistical mistakes that

went unnoticed by the lower court and the first court of appeal and these resulted in the wrongful conviction of Mrs. Clark. In an official statement, the Royal Statistical Society said: “The case of *R v. Sally Clark* is one example of a medical expert witness making a serious statistical error, one which may have had a profound effect on the outcome of the case. [...] The Society urges the Courts to ensure that statistical evidence is presented only by appropriately qualified statistical experts” (2001). The advice of the Society may be one step in the right direction, but it does not seem enough to prevent courts from making mistakes. Even if statistical evidence is only presented by qualified experts, judges and other legal factfinders still face the problem of correctly interpreting and applying these findings in their evidential decision-making.

In this chapter, I discuss the risks of the incorrect use of probabilities in court and the question of what to do about them. I examine the nature of these risks and the intricate interplay between risks and responsibilities within the rule of law. For practical reasons, I restrict myself to a discussion of criminal law and I take most of my examples from Dutch criminal law because it is the system that I am familiar with. However, my analysis is meant to be relevant for other legal systems, in particular for continental systems in which judges, not juries, are the factfinders.

In the next section, I first briefly discuss an experiment to show how the use of probability theory in court can easily result in fallacious judicial evidential reasoning. Subsequently, I analyze the nature of the risks of misinterpretation and misapplication of probabilistic findings in more detail. I also discuss the question who should be responsible for reducing those risks. Next, a large part of this chapter is then devoted to a discussion of three possible solutions to reduce the risk of probabilistic errors in court. I pay close attention to the demand that all solutions must be in accordance with the rule of law in general and with the demands of a fair trial in particular. The first and most evident solution is more judicial training in probability theory. However, this does not seem sufficient to reduce the risk of errors, or so I shall argue. The second possible solution is the introduction of probability experts as what I call “probability clerks.” This solution seems to suffice in evidentially simple cases, but not in evidentially complex cases. I argue that we need a third and more radical solution, viz. the introduction of what I call “probability judges”, at least in evidentially complex cases.

My main conclusions are that we do not know the number and the severity of probabilistic errors in court, but that we have reasons to worry about them, in particular about the risks of miscarriages of justice. My main recommendations are, first, that empirical research be done to investigate the precise number and the nature of the risks and, second, that experiments with probability clerks and probability judges be done in order

to empirically test whether their introduction can reduce the number and severity of probabilistic fallacies in court.

## 5.2 Probabilistic Reasoning in Court: An Example

I start with an example that illustrates the role that probabilistic reasoning nowadays plays in forensic reports. It offers insight into one type of probabilistic misunderstanding that these forensic reports can cause in judicial evidential decision-making. The example is a simplified criminal case of a robbery at a cash dispenser which I take from an experiment by De Keijser and Elffers (2012, 195–8). The two main pieces of evidence are security camera images of the robber and a report of a forensic expert who compared these images with photos of the suspect.

The expert reports that he has carried out comparative research and that he has examined if the findings fit better under hypothesis 1 than under the alternative hypothesis 2. Hypothesis 1 holds that the suspect is the perpetrator (more specifically, it holds that the perpetrator of the robbery visible on a specific CCTV image is the same person as the suspect depicted in a specific photo). Hypothesis 2 holds that the suspect is not the perpetrator (more specifically, it holds that the perpetrator of the robbery visible on the CCTV image is not the same person as the suspect depicted in the photo). The expert reports that the findings based on the selected visual materials of the facial comparison are much more likely when the person depicted is the same person (hypothesis 1) than when they are different persons (hypothesis 2).

De Keijser and Elffers asked judges, defense lawyers and experts what they can correctly derive from this report. Among others they asked the participants whether the following is a correct interpretation of the conclusion of the expert: “It is much more likely that the suspect is the person on the images from the security camera than someone else is the person on those security camera images” (De Keijser and Elffers 2012, 198). More than 88% of the judges and lawyers and more than 63% of the experts believed this conclusion to be correct (De Keijser and Elffers 2012, 199–200).<sup>3</sup> Unfortunately, however, it is false. The expert reports on the probability that one will find the evidence, *given a particular hypothesis*, but most participants – including a majority of the forensic experts – interpret the statement as a report on the probability that a particular hypothesis is true, *given the evidence*. The mixing up of these probabilities is called the prosecutor’s fallacy (Thompson and Shumann 1987).<sup>4</sup> An even more worrisome finding of De Keijser and Elffers is that more than half of the judges and defense lawyers and 85% of the experts claimed to have a perfect or near-perfect understanding of the forensic conclusions presented to them (2012, 201–2). In other words, not only did a majority of the participants

misinterpret the report, but many of them were also blind to their own lack of understanding.

One needs to have basic knowledge of Bayesian probability theory to understand the prosecutor's fallacy. For the purposes of this chapter, however, there is no need to go into the details of probability theory. It suffices if the reader has an intuitive grasp of the nature and of the potential far-reaching consequence of this type of mistake. Let me therefore present a simple example. Compare the following two questions. First, what is the conditional probability that a randomly chosen mammal has *four legs*, if (condition) it is a *cow*? The probability that a randomly chosen cow has four legs is quite high. Second, what is the conditional probability that a randomly chosen mammal is a cow, if (condition) it has *four legs*? This probability seems very low. This example helps to understand that these two conditional probabilities can diverge dramatically. Now we see more clearly that it is one thing to say that there will probably be a match between the findings if the defendant is the perpetrator, but quite a different claim to say that the defendant is probably the perpetrator if there is a match. If courts mix up these probabilities, like the participants of the experiment of De Keijser and Elffers did, they run the risk of reaching incorrect conclusions about the probability that the defendant committed the crime and therewith they run the risk of committing a miscarriage of justice.

### 5.3 Responsibility and Risk

I began this chapter by referring to the risks of incorrect use of probabilities in court and the question of what to do about them. We have just seen that mistakes in the interpretation and application of probabilistic statements can result in fallacious argumentation, false conclusions, and – in the worst case – miscarriages of justice.

Risk is standardly defined as the statistical expectation value of an unwanted event that may or may not occur, or as the product of the probability that an event will take place and the degree of “unwantedness” or severity of that event (Hansson 2018). Accordingly, to say that a risk is high can mean that both the probability and the severity of an event are high, that the probability of the event is high but the severity low, or that the probability is low but the severity high. However, if the severity is deemed very low, we no longer speak in terms of risks.

#### 5.3.1 *First-Order Risk and Second-Order Risk*

The experiment of De Keijser and Elffers and the other literature I referred to suggest that the probability that judges and other legal factfinders make mistakes in probabilistic reasoning is quite high. This holds in particular

for the prosecutor's fallacy. Next to that, several other probabilistic fallacies have been distinguished in the literature. For example, Dahlman (2018) has distinguished ten types of probabilistic errors, such as base rate neglect, underestimating the combined strength of concurring evidence and dependence neglect.<sup>5</sup>

However, as far as I know, we do not know how often courts make probability mistakes, nor do we know how severe the consequences of these mistakes are. There are indications that the prosecutor's fallacy is made regularly, and the same seems to hold for the base rate neglect and for the underestimation of the combined strength of weak evidence. However, even if we assume that courts regularly make these mistakes, then we still do not know how often these fallacious inferences result in miscarriages of justice.

In other words, we are not only confronted with first-order risks, that is, with the risk of probability mistakes in court and the risk that these mistakes result in miscarriages of justice. Because we are ignorant both about the probability of probability mistakes and about the severity of their consequences, we are uncertain about the magnitude of the first-order risk. Therefore, we are also confronted with second-order risks.<sup>6</sup> If we incorrectly believe the first-order risk is high, we will spend too much effort on preventing mistakes. Conversely, if we underestimate the first-order risk, we end up taking insufficient preventive measures. In conclusion, we need empirical research to determine the first-order risk of judicial probability mistakes and therewith to lower the second-order risk.<sup>7</sup>

### 5.3.2 *Risk versus Uncertainty; Objective versus Subjective Probability*

Given that we are uncertain about the number as well as about the severity of the unwanted effects of probability mistakes, some readers might want to object to my use of the term "risk." Following Knight's distinction between risk and uncertainty, they could argue that for lack of quantifiable probabilities about probability mistakes, I should speak in terms of uncertainty. Knight states: "To preserve the distinction [...] between the measurable uncertainty and an unmeasurable one we may use the term 'risk' to designate the former and the term 'uncertainty' for the latter. [...] We can also employ the terms 'objective' and 'subjective' probability to designate the risk and uncertainty respectively, as these expressions are already in general use with a signification akin to that proposed" (Knight 1921, 233).

Let me make two brief remarks on this issue. First, the probabilities that we are interested in are as yet unknown, but they are not unknowable in principle. Empirical research could deliver the statistical information we need and, in fact, one of my recommendations is that research be done to

gather that information. Second, as the quote makes clear, Knight does not only use the terms risk and uncertainty, but he also refers to the distinction between objective and subjective probability. Let me explain the difference between these two types of probability (Hacking 2001, 132–7).<sup>8</sup> If a coin has been tossed 100 times and landed heads 45 times, then the frequency of heads of this coin is .45. If we say that (in the long run) the probability of getting heads with this coin is .45, we seem to be using an objective or frequency-type probability. However, if we want to assess the probability that the coin will land heads the next time I toss it, talking in terms of frequency-type probability does not make sense. On a single occasion, a coin will either land head or tails. In a single case, we can only use a subjective or belief-type probability. We can say that our degree of belief that the coin will land heads is .45, even though the reason to have this degree of belief is the information I have about the frequency that the coin has landed heads in the past. The same holds for the example about the probability that the CCTV images and the photo of the defendant match if the defendant is the perpetrator. This too is a belief-type probability. Judicial decision-making is about single cases. Therefore, judicial decision-making is about belief-type, that is, Bayesian, probabilities. Therefore, in judicial decision-making, the distinction between uncertainty and risk is not a fundamental or principled distinction, because on a Bayesian account, all probabilities, even those informed by “objective” frequencies, are subjective.

### 5.3.3 *Material Risks and Epistemic Risks*

Let us return to the risk of probability mistakes in judicial decision-making. I have argued that we are uncertain about the probability and the severity of probability mistakes made by courts. However, we can be more precise about the nature of the risks involved in the judicial interpretation and application of probabilistic statements. We can distinguish two types of risk. First, there is the risk that courts make unsound inferences and that they, as a consequence of these unsound inferences, come to adhere to false beliefs. Second, these false beliefs can have further adverse consequences, incorrect decisions – in the worst case miscarriages of justice – and their executions.

When we talk about risks, we often focus on the latter type of risk, i.e., on the material or practical risk of (the execution of) wrongful convictions and wrongful acquittals that can follow from unsound inferences and false beliefs. The reason to call them material or practical risks is that they put our value of practical rationality at risk because and to the extent that they are about making and executing or implementing legally, morally, and politically wrongful decisions. However, the events of making

unsound inferences and adhering to false beliefs are not only unwanted because of their undesirable practical consequences. Committing fallacies and adhering to false beliefs are also in themselves unwanted, because and to the extent that they conflict with our value of epistemic or theoretical rationality.<sup>9</sup> Epistemic rationality is not only a scientific value, but also a fundamental value of criminal trials since these trials aim not merely at a procedural truth, but primarily at the material truth. Accordingly, if a court, by moral luck, makes a correct legal decision that is based on flawed reasoning and/or false beliefs, epistemic injustice is nevertheless done. Therefore, the risks of committing fallacies and adhering to false beliefs are called epistemic risks.

Accordingly, we can distinguish between practical or material risks of wrongful decisions and their execution on the one hand and theoretical or epistemic risks on the other. At least three different types of epistemic risk play a role in the judicial interpretation and application of probability statements in legal cases. The first epistemic risk is the second-order risk mentioned above, i.e., the fact that we are deeply uncertain about the nature and the magnitude of the first-order risk, viz. about the probability and the severity of courts making probability mistakes. The second epistemic risk is the risk of making unsound inferences, regardless of whether these inferences result in false beliefs. The third epistemic risk is the risk of actually entertaining false beliefs as a consequence of unsound reasoning.

#### 5.3.4 *Who Should Be Responsible?*

Before turning to the question whether it is possible to reduce these risks, let us briefly discuss the question of who is or should be responsible for assessing and reducing them. At first sight, it seems quite logical to say that the judiciary as a state organ is responsible for assessing the quality and the quantity of the risks of making probability mistakes in judicial evidential reasoning and that both the judiciary as an organization and individual judges are responsible for minimizing these risks. Both the judiciary as a whole and individual judge must ensure that courts are competent to perform their task of evidential decision-making. Like other professionals, they need to see to their own training and ask for advice if they lack specific expertise. Accordingly, if they lack competence in forensic sciences and probability theory, they should get more training and/or advice.

However, the judiciary and judges functioning in the rule of law differ from other professionals and professional organizations in some crucial respects. If professionals lack competence for a particular task, either they will not perform the task themselves and refer clients to another



professional who is competent, or they will collaborate in a team of professionals so as to make sure the team as a whole has the required competence. Judges, however, cannot operate in a similar way. First, they are not allowed to refuse to decide a case; that would be a denial of justice.<sup>10</sup> Second, to ensure their independence and impartiality, judges are not allowed to collaborate with other professionals when deciding a case. Courts can ask experts for advice, but they have to make the decision on their own: experts are allowed, as advisors, in the court room but not, as decision makers, in the council chamber. For the same reason, even though courts can discuss general characteristics of a case with “outsiders” and ask for general advice, they are not allowed to discuss their envisaged decisions in specific cases.

These restrictions have ramifications for the nature of possible solutions. In the first place, the judiciary can only take measures within the confines of the law. Secondly, although the legislator can change the law, it must see to it that the solutions are in accordance with fundamental human rights, in particular with the right to a fair trial as it is laid down in constitutions and in international treaties like article 6 of the European Convention on Human Rights (ECHR) and article 14 of the International Covenant on Civil and Political Rights.

#### 5.4 Three Possible Solutions

In this section, I discuss three possible solutions that are in accordance with the rule of law and the right to a fair trial. The first solution does not demand any adaption of (Dutch) criminal law, the second demands a slight adaptation, and the last is the most revisionary proposal. As far as I can see, they are the only feasible solutions within the bounds of the rule of law and the (Dutch) system of law.

##### 5.4.1 *Training*

The first and most obvious solution to reduce the number of probability mistakes is to provide judges with more education in probability theory. Through these trainings, judges can acquire passive understanding of probability theory, in particular of Bayes’ rule, and of important concepts such as the prior probability and the likelihood ratio. However, it seems much more difficult and possibly too time-consuming to acquire the ability to reason actively and correctly with probabilities and to detect errors in probability reasoning of oneself and others. Teaching judges basic understanding of probability theory is definitely necessary, but it does not seem sufficient to prevent them from making serious probabilistic mistakes.<sup>11</sup>

#### 5.4.2 *Probability Clerks*

Another possible solution is the appointment of probability clerks, i.e., assistants with specific expertise in probability theory. This solution fits nicely with recent developments in the Dutch judiciary. In 2012–13, a Forensic Support pilot has been conducted in a number of courts. The pilot consisted of the appointment of forensic assistants, generalists with a master's degree in forensic sciences (Raad voor de rechtspraak 2014, 5). In 2014, the pilot had been evaluated positively, and it resulted in the appointment of forensic assistants at all criminal courts, both lower courts and courts of appeal (Raad voor de rechtspraak 2014).<sup>12</sup>

The task of the forensic assistants is, among others, to prepare the forensic parts of criminal files and to answer clarificatory questions from judges about forensic reports and about the hearings of forensic experts. By analogy, assistants who have obtained a master's degree in probability theory or statistics could be appointed as probability clerks. The advantage of appointing probability experts as assistants next to expert witnesses is that probability clerks can explicate probability arguments in the forensic reports, not only in the preparation for the hearing, but also after the hearing and in the council chamber.

Of course, we do not know whether the introduction of probability clerks would be as successful as the introduction of forensic assistants. A pilot should be conducted to find out whether probability clerks can help to reduce the number and the severity of probability mistakes made by courts. However, there are several reasons why we should doubt that their appointment suffices in evidentially complex cases. I mention three limitations in particular. First, probability clerks can prepare questions for the court, but as clerks they are not allowed to ask (follow-up) questions during trial at the hearing of the experts. At the crucial moment of the hearing, the judge has to ask the proper questions without assistance. Second, although probability clerks can, like forensic assistants, be present in the council chamber to answer clarificatory questions, they are not allowed to participate in the deliberations. Moreover, and this is a third limitation, as clerks, they are not allowed to give their own interpretations of probability statements because they are not experts in the sense of the law (Raad voor de rechtspraak 2014, 5).<sup>13</sup>

#### 5.4.3 *Probability Judges*

The limitations on the role of probability clerks suggest that we need probability judges, at least in evidentially complex cases.<sup>14</sup> The introduction of probability judges might sound problematic. Readers might worry that it conflicts with the rule of law and with the continental view that jurists have a monopoly on the judiciary and that they have it for good reasons.

A third possible objection is that my plea for probability judges opens the floodgates to the introduction of many more types of expert judges. Let me discuss these three worries in turn.

*5.4.3.1 The Monopoly of Legal Professionals in the Judiciary*

First, it should be noted, that even in continental countries the juridical monopoly on the judiciary has never been complete. Most countries have lay judges and/or some mixed chambers of the court. Second and more importantly, we can observe an analogy between the current situation and the situation at the beginning of the 19<sup>th</sup> century when legal professionals obtained the monopoly on the judiciary. By the end of the 18<sup>th</sup> century, law had become so complex in many Western countries that it was no longer deemed sufficient that courts consisted of lay judges who let themselves be advised by a legal professional. The turning point in the Netherlands was the Dutch Code on the Judicial Organization of 1827 that ordained that all judges of lower courts, courts of appeal, and the supreme court should be jurists with specific legal training and competence (Van Boven 1990, 267–70).

The analogy between the 1820s and the 2020s can easily be seen. In those days, it was the law; in our days, it is not only the law, but also the assessment of evidence, especially of forensic evidence that has become too complex to be handled by lay persons. The interpretation and application of evidential findings has become so complex that it demands specific probabilistic competence. As in the 1820s, there are reasons to believe it is no longer sufficient that the judge is advised by experts, but that probability experts should themselves be members of the court.

*5.4.3.2 The Right to a Fair Trial*

This takes us to the question whether probability judges are in accordance with the role of the judiciary in the rule of law. The fundamental task of the judiciary lies not only with the correct application of material law, but also or even primarily with safeguarding a procedure that ensures a fair trial. Article 6 (1) ECHR, for example, states that “In the determination of his civil rights and obligations or of any criminal charge against him, everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law.” The introduction of probability judges does not seem to conflict with the demands of a fair, public, and timely hearing or with the independence and impartiality of the court. Moreover, another crucial aspect of a fair trial is that, in the end, the court delivers an understandable and properly reasoned judgment. As I have argued in foregoing sections, this is exactly what is at stake in evidentially complex cases. Stated yet differently, we

need probability judges to fulfill a fundamental requirement of a fair trial, viz., that courts deliver properly reasoned judgments.

Another important aspect of a fair trial is the adversarial principle. One aspect of this principle is the requirement that parties have had sufficient opportunity to react to all the evidence and the arguments. This implies that judges cannot discuss any insights, reasoning, or information in the council chamber or present them in their final decision if they have not been discussed during the hearing. It is the task of the legally trained presiding judge to guard this and other important aspects of a fair trial and the probability judge should be trained to act in accordance with this fundamental adversarial principle.

#### 5.4.3.3 *Opening the Floodgates?*

A final objection to my plea for probability judges is that it opens the floodgates to the introduction of many more types of expert judges. However, I believe that this objection fails too. First, it should be noted that the expertise of probability experts differs substantially from the expertise of other experts. For one thing, probability experts do not have, like forensic experts, “substantive” expertise about the material or “underlying” facts. Probability theorists are experts with respect to reasoning with and about probabilities. Since all forensic experts make probabilistic claims, probability judges can assess the quality of their probabilistic arguments. In this respect, probability experts are on a par with legally trained judges who also lack substantive expertise and who are “only” experts in reasoning with and about legal rules. In seeing to it that cases are decided in accordance with legal rules, legally trained judges are the guardians of practical rationality (Schauer 1993). Similarly, in seeing to it that cases are decided in accordance with the rules of probability theory, probability judges would be the custodians of theoretical rationality.

Therefore, or so I conclude, my proposal does not open the floodgates to a whole array of other types of expert judges. On the contrary, given that probability judges can strengthen both the critical evaluation of forensic reports and the evidential reasoning of the court itself, my proposal is perhaps the best way to manage and improve the ever-increasing contributions of forensic experts in the administration of justice, at least in evidentially complex cases.

## 5.5 **Conclusions and Recommendations**

In this chapter, I have discussed the risks involved in the judicial interpretation and application of probability statements. I have argued that there are material risks and three types of epistemic risks. First, for lack of solid

empirical research into the matter, we are uncertain both about the probability and about the severity of probability mistakes made by criminal courts (second-order epistemic risk). However, there are several indications that the probability of judicial probability mistakes is quite high. These mistakes consist in “unsound”<sup>15</sup> inferences such as the prosecutor’s fallacy (first-order epistemic risk) and they often result in false beliefs (another first-order epistemic risk). The ultimate risk is that they result in wrongful convictions and acquittals and their executions (first-order material risks). There are examples of wrongful convictions, such as the infamous case of *R v. Sally Clark*, but again we seem to lack reliable numbers.

Even though we do not know the number and the nature of judicial probability mistakes nor the number and nature of their adverse practical consequences, there are indications that judges are insufficiently competent in interpreting and applying probabilistic statements. Therefore, more training of judges is necessary. I have argued, however, that such training does not seem sufficient. Even though it is possible to teach judges basic but fairly passive knowledge of probability theory, it seems much harder and more time-consuming to teach them how to actively use probability theory themselves, to critically question experts, and to detect flaws in probabilistic reasoning.<sup>16</sup>

Therefore, I have argued for more far-reaching changes of our legal systems. One relatively simple change is the introduction of experts in probability theory as clerks who can explicate probabilistic statements of experts and who can help to prepare questions for experts and who can help judges to avoid making probability mistakes in their evidential argumentation. On the positive side, the introduction of probability clerks fits easily in existing legal systems. In fact, it is just one step beyond the recent successful introduction of forensic assistants in Dutch criminal courts. On the negative side, I have argued that the introduction of probability clerks might not be sufficient as a solution in evidentially complex cases.

My third and most radical proposal has been the introduction of probability judges, i.e., experts in probability theory who sit themselves as judges in mixed chambers of the court. I have argued that their introduction seems necessary in evidentially complex cases and I have argued their introduction is possible within the confines of the rule of law and without running the risk of opening the floodgates to many other types of expert judges.

Finally, I call for two types of empirical research. First, I call for research into the nature and the number of probability mistakes that are being made by criminal courts and into their material consequences. Second, I call for experiments with probability clerks and probability judges to empirically test my hypothesis that their introduction contributes to reducing the number and the severity of probability mistakes made by courts.

## Notes

- 1 This is sometimes called the advisory system of the judiciary. The alternative is a decision system in which judicial professionals decide cases together with professionals from other disciplines. In Western legal systems, the advisory system is the standard and the decision system the exception. One of the Dutch exceptions to the advisory system is the penitentiary chamber of the Court of Appeal Arnhem-Leeuwarden. This chamber consists of three judges and two behavioral experts (a psychiatrist and a psychologist). See De Groot and Elbers (2008).
- 2 See, for example, Lagnado (2021).
- 3 Perhaps the most surprising fact is that the mistakes were not only made by a majority of the judges and lawyers, but even by a majority of the forensic experts. Some forensic experts have or at least are assumed to have knowledge of Bayesian probability theory. However, it should be noted that this depends on the discipline of the experts. Not all forensic experts are trained to apply and to report in terms of Bayesian probability theory.
- 4 There are some indications that the prosecutor's fallacy is frequently committed by Dutch Courts. Prakken (2018) analyzed 31 recent Dutch judicial decisions and found that the court committed the prosecutor's fallacy in 22 of them. Also see Prakken and Meester (2017). Meester and Stevens (2021) analyzed another four recent Dutch criminal cases and detected the prosecutor's fallacy, the base rate neglect, and the underestimation of the combined strength of weak evidence in them.
- 5 Dahlman (2018) further distinguishes false positive neglect; wrong reference class; false dichotomy; underestimating the cumulative uncertainty in evidence chains; double-counting and double-discounting; overestimating predictive evidence. Also, see Dahlman, in preparation.
- 6 See Möller, Hansson, and Peterson (2006, 422ff) about second-order risks and epistemic uncertainty.
- 7 These are not the only uncertainties, however. Human beings do not only err in probabilistic reasoning, but also in many other ways and we do not know whether the risks of fallacious probabilistic reasoning are higher than the risks of other kinds of fallacies and biases. Finally, we also do not know whether debiasing measures are effective (Zenker 2021).
- 8 For an application in law see, for example, Robertson and Vignaux (1993).
- 9 On the distinction and the relationship between practical and theoretical rationality, see Mackor (2011) and Mackor (2013).
- 10 In Dutch law, this prohibition is laid down in article 13 Wet Algemene Bepalingen [General Provisions Act].
- 11 Moreover, elsewhere I have argued extensively that judges not only need to learn probability theory, but also explanation-based theories to evidential decision-making such as the scenario theory (Mackor and Van Koppen 2021; Mackor, Jellema, and Van Koppen 2021).
- 12 And through personal correspondence with the Rechtspraak Servicecentrum [Service Centre of the Judiciary] July 22, 2020.
- 13 The Evaluatie Pilot Forensische ondersteuning rechtbanken Straf commissioned by the Dutch Council for the Judiciary explicitly states that forensic assistants are not experts in the sense of the Code of Criminal Procedure and that they do not provide their own interpretation of the forensic evidence or the criminal case. The same would be true of probability clerks.
- 14 For a more detailed analysis of the question whether and how the introduction of probability judges in the Netherlands is possible within the limits of the Dutch constitution, see Mackor and Schutgens (2022).

- 15 Here, I define unsafe or unsound convictions and acquittals as decisions that are in themselves correct but based on unsound reasoning. Stated differently, in such cases, the conviction or the acquittal can be upheld, but not on the basis of the argumentation of the verdict. I distinguish these cases from wrongful convictions and acquittals in which the decision cannot be upheld.
- 16 In 2022, Mackor, Dahlman, and Lagnado received a NWO research grant (number 406.21.RB.004) to develop and test a method for teaching judges to reason more rationally about evidence in criminal cases. More information at <https://preventingmiscarriagesofjustice.wordpress.com/>

## References

- Dahlman, Christian. 2018. *Beviskraft. Metod för bevisvärdering i brottmål*. Stockholm: Norstedts.
- Dahlman, Christian. In preparation. A Systematic Account of Probabilistic Fallacies in Legal Fact-finding.
- De Groot, Dineke, and Nieke Elbers. 2008. *Inschakeling van deskundigen in de rechtspraak: Verslag van een onderzoek naar knelpunten en verbetervoorstellen*. Research Memoranda 4(3). The Hague: Raad voor de rechtspraak. <https://www.rechtspraak.nl/SiteCollectionDocuments/Inschakeling-van-deskundigen-in-de-rechtspraak.pdf>
- De Keijser, Jan, and Henk Elffers. 2012. "Understanding of Forensic Expert Reports by Judges, Defense Lawyers and Forensic Professionals." *Psychology, Crime & Law* 18 (2): 191–207.
- Hacking, Ian. 2001. *An Introduction to Probability and Inductive Logic*. Cambridge: Cambridge University Press.
- Hansson, Sven Ove. 2018. "Risk." In *The Stanford Encyclopedia of Philosophy (Fall 2018 ed.)*, edited by Edward N. Zalta. Stanford: Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/fall2018/entries/risk/>
- Knight, Frank H. 1921. *Risk, Uncertainty, and Profit*. Boston: Houghton Mifflin Company.
- Lagnado, David. 2021. *Explaining the Evidence: How the Mind Investigates the World*. Cambridge: Cambridge University Press.
- Mackor, Anne Ruth. 2011. "Explanatory Non-Normative Legal Doctrine. Taking the Distinction between Theoretical and Practical Reason Seriously." In *Methodologies of Legal Research. Which Kind of Method for What Kind of Discipline*, edited by Mark Van Hoecke, 45–70. Oxford: Hart Publishing.
- Mackor, Anne Ruth. 2013. "What Can Neurosciences Say about Responsibility? Taking the Distinction between Theoretical and Practical Reason Seriously." In *Neuroscience and Legal Responsibility*, edited by Nicole A. Vincent, 53–83. Oxford: Oxford University Press.
- Mackor, Anne Ruth, Hylke Jellema, and Peter J. Van Koppen. 2021. "Explanation-Based Approaches to Reasoning about Evidence and Proof in Criminal Trials." In *Law and Mind: A Survey of Law and the Cognitive Sciences*, edited by Bartosz Brozek, Jaap Hage and Nicole A. Vincent, 431–70. Cambridge: Cambridge University Press.
- Mackor, Anne Ruth, and Roel Schutgens. 2022. Artikel 116 Grondwet: de inrichting en samenstelling van de rechterlijke macht. Over het monopolie van juristen en de positie van de kansdeskundige. In *Een nieuw commentaar op de Grondwet*, edited by Afshin Ellian and Bastiaan Rijpkema, 451–66. Amsterdam: Boom.

- Mackor, Anne Ruth, and Peter J. Van Koppen. 2021. "The Scenario Theory about Evidence in Criminal Law." In *Philosophical Foundations of Evidence Law*, edited by Christian Dahlman, Alex Stein and Giovanni Tuzet, 213–24. Oxford: Oxford University Press.
- Meester, Ronald, and Lonke Stevens. 2021. "Correct redeneren: wat een Bayesiaanse analyse zegt over vier recente strafrechtelijke uitspraken." *Expertise en Recht* 14 (3): 112–21.
- Möller, Niklas, Sven Ove Hansson, and Martin Peterson. 2006. "Safety Is More than the Antonym of Risk." *Journal of Applied Philosophy* 23 (4): 419–32.
- Prakken, Henry. 2018. "Kansoordelen door deskundigen: over 'logisch' rapporteren en wat daarbij mis kan gaan." *Ars Aequi* 67: 740–7.
- Prakken, Henry, and Ronald Meester. 2017. "Bayesiaanse analyses van complexe strafzaken door deskundigen. Betrouwbaar en zo ja: Nuttig?" *Expertise en Recht* 10 (5): 185–97.
- Raad voor de rechtspraak. 2014. *Evaluatie Pilot Forensische ondersteuning rechtbanken Straf*. The Hague: Raad voor de rechtspraak. [https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Nieuws/Documents/\(256140063\)%20EVALUATIE%20PILOT%20FORENSISCHE%20ONDERSTEUNING%20\).pdf](https://www.rechtspraak.nl/Organisatie-en-contact/Organisatie/Raad-voor-de-rechtspraak/Nieuws/Documents/(256140063)%20EVALUATIE%20PILOT%20FORENSISCHE%20ONDERSTEUNING%20).pdf)
- Robertson, Bernard, and GA Vignaux. 1993. "Probability. The Logic of the Law." *Oxford Journal of Legal Studies* 13 (4): 457–78.
- Royal Statistical Society. *Royal Statistical Society Concerned by Issues Raised in Sally Clark case*, 23 October 2001. <https://rss.org.uk/RSS/media/File-library/Membership/Sections/2020/Sally-Clark-RSS-statement-2001.pdf>
- Schauer, Frederick. 1993. *Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life*. Oxford: Clarendon Press.
- Thompson, William C., and Edward L. Shumann. 1987. "Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy." *Law and Human Behavior* 2 (3): 167–87.
- Van Boven, Maarten Willem. 1990. *De rechterlijke instellingen ter discussie. de geschiedenis van de wetgeving op de rechterlijke organisatie in de periode 1795-1811*. PhD thesis. Nijmegen.
- Zenker, Frank. 2021. "Debiasing Legal Fact-Finders." In *Philosophical Foundations of Evidence Law*, edited by Christian Dahlman, Alex Stein and Giovanni Tuzet, 395–409. Oxford: Oxford University Press.