Perspectives in
Legal and Forensic Sciences

SCIEPublish

Article

Statistical Theatre: Misinterpretation of Quantitative Evidence in the Courtroom

John S. Croucher [1,*] and Rebecca C. K. Ward [2]

[1] Department of Actuarial Studies and Business Analytics, Macquarie Business School, Macquarie University, Sydney, NSW 2109, Australia

[2] College of Business, Law and Governance, James Cook University, Townsville, QLD 4811, Australia

* Corresponding author. E-mail: john.croucher@mq.edu.au (J.S.C.)

Received: 14 April 2026; Revised: 26 May 2026; Accepted: 6 July 2026; Available online: 24 July 2026

ABSTRACT: The increasing use of quantitative evidence in legal proceedings reflects a broader shift towards data-informed forms of proof. Statistical analyses, probability estimates, and forensic calculations are frequently presented as objective indicators of truth; however, their evidential value depends not on the mathematics itself, but on how the relationships they describe are interpreted. This paper examines the misinterpretation of quantitative evidence in courtroom settings, arguing that numerical outputs are often treated as conclusions rather than as components of structured inference. Focusing on conditional probability, the prosecutor's fallacy, base rate neglect, witness testimony, and DNA evidence, the paper demonstrates how common errors arise from a failure to engage with the conditional and relational nature of probabilistic reasoning. Consistent with earlier work highlighting the interpretive limits of quantitative evidence, the analysis of key cases, including the Sally Clark case and People v Collins, together with contemporary examples drawn from forensic science and algorithmic decision-making, demonstrates how numerical evidence can assume persuasive authority that exceeds its probative value when underlying assumptions are not made explicit. Building on established scholarship concerning the persuasive authority of numerical evidence, expert testimony, and probabilistic reasoning in legal decision-making, this paper proposes the Statistical Theatre Model to describe situations in which quantitative evidence acquires persuasive force independent of its inferential value. The paper further considers cognitive and institutional factors that contribute to these errors and argues that improvement lies not in increased mathematical complexity, but in greater conceptual clarity. In addition to identifying common interpretive failures, the paper proposes practical reforms to improve the communication and evaluation of quantitative evidence by experts, lawyers, judges, and jurors. In doing so, it highlights the importance of aligning the presentation of quantitative evidence with the interpretive demands of legal decision-making. Statistical evidence must remain a tool of inference rather than an unwarranted source of certainty.

Keywords: Statistical evidence; Conditional probability; Prosecutor's fallacy; Base rate neglect; DNA evidence; Witness testimony; Bayesian reasoning; Legal decision-making; Statistical independence; Likelihood ratios

1. Introduction

The increasing reliance on quantitative evidence in legal proceedings reflects a broader shift towards scientific and data-informed forms of proof. Statistical analyses, probability estimates, and forensic calculations are now routinely presented in courtrooms as objective indicators of truth [1,2]. Yet, despite their apparent precision, such forms of evidence are frequently misunderstood. The issue is not the mathematics itself, but the interpretation of the relationships that the mathematics describes. This interpretive gap has become more pronounced in contemporary legal contexts, where increasingly complex forms of evidence—including probabilistic modelling, forensic genomics, and algorithmic decision-making—are introduced into courtrooms that are not always equipped to evaluate them appropriately [3,4]. Recent work in both legal technology and forensic practice further suggests that this gap is widening rather than resolving, as the complexity of evidential forms outpaces the development of shared interpretive frameworks [5,6].

This paper focuses primarily on common law adversarial jurisdictions, including Australia, the United Kingdom, Canada, and the United States. Although many of the issues discussed may arise in other legal traditions, the analysis is directed towards systems in which expert evidence is presented within an adversarial framework and evaluated by judges and juries.

As Devlin [7] observes, mathematics is fundamentally concerned with relationships rather than numbers alone. Statistical reasoning, in particular, requires an understanding of how quantities relate to one another under specific conditions [1,8]. For the purposes of this paper, statistical reasoning is used as a broad term encompassing probabilistic reasoning, Bayesian reasoning, likelihood-based reasoning, and related inferential approaches that seek to evaluate evidential relationships under conditions of uncertainty. Numbers do not speak for themselves; they derive meaning only within a framework of assumptions, context, and inference. Statistical outputs are therefore not conclusions in themselves, but representations of conditional relationships. When these relationships are overlooked, numerical outputs can be misinterpreted, sometimes with significant legal consequences [9].

The consequences of such misunderstandings extend beyond technical error. Misinterpretation of statistical evidence has been implicated in wrongful convictions, flawed expert testimony, and distorted assessments of risk and reliability [8]. High-profile cases such as Sally Clark and *People v Collins* remain widely cited examples, while more recent controversies involving probabilistic genotyping, algorithmic risk assessment, and forensic decision-support systems demonstrate that these problems remain contemporary rather than historical.

One of the most persistent difficulties lies in the interpretation of probabilistic evidence. Courts are frequently required to evaluate statements about likelihoods—whether expressed in terms of match probabilities, error rates, frequencies, or likelihood ratios—yet these are often treated as direct indicators of guilt or innocence [10]. This reflects a fundamental misunderstanding. A probability associated with evidence is not equivalent to a probability of a hypothesis. This distinction can be understood as the difference between the probability of observing the evidence if a hypothesis were true and the probability that the hypothesis is true given the observed evidence. These are not interchangeable and can yield vastly different answers. Confusion between these concepts, commonly referred to as the prosecutor's fallacy, has been widely documented [9,11]. Importantly, in many forensic disciplines that operate within a Bayesian framework, experts are expected to express the evidential value of observations through likelihood ratios rather than opinions on posterior probabilities or ultimate issues of guilt. Nevertheless, research suggests that likelihood ratios may still be misunderstood by legal decision-makers and interpreted as direct indicators of guilt or innocence despite repeated efforts to distinguish the two [1,12].

Despite decades of scholarly attention, these errors persist. Research continues to identify difficulties in the communication of probabilistic evidence, while legal decision-makers remain vulnerable to inferential errors involving conditional probability, base rate neglect, and statistical independence [5,11]. Experimental research further suggests that such errors may persist even when the relevant statistical distinctions are explicitly explained, indicating that the problem extends beyond technical knowledge to broader issues of evidential communication and cognitive processing [12–14].

Building on this literature, this paper proposes the Statistical Theatre Model. The term is used to describe situations in which quantitative evidence acquires persuasive authority independent of its actual inferential value. The underlying phenomenon is not new. Previous scholarship has examined the persuasive influence of numerical evidence, probabilistic reasoning, expert testimony, and evidential communication within legal settings [10,14,15]. The contribution of the present paper is therefore not the identification of a previously unrecognised phenomenon, but the development of an integrative model that synthesises these recurring interpretive failures within a common explanatory framework. Under conditions of uncertainty, numerical outputs may create an appearance of objectivity and certainty that exceeds what the underlying reasoning can justify. The result is not necessarily fraudulent or misleading evidence, but evidence that is granted persuasive force because of its numerical form rather than its evidential strength.

The Statistical Theatre Model is presented later in the paper after the principal case studies and analytical examples have been examined. This structure is intentional. Rather than imposing a theoretical framework at the outset, the paper first identifies recurring patterns across diverse examples before drawing those observations together within a unified explanatory model. The model, therefore, functions as a synthesis of the preceding analysis rather than as a starting assumption.

This paper examines the ways in which quantitative evidence is misinterpreted in the courtroom, with particular attention to conditional probability, statistical independence, base rate neglect, and the evaluation of witness and forensic evidence. Drawing on established statistical principles [1,8], Bayesian and likelihood-based approaches to evidential reasoning [11,12], and recent interdisciplinary developments across forensic science, medicine, legal technology, and contemporary scholarship on probabilistic reasoning, it argues that many of these errors arise from treating numbers as conclusions rather than as components of a broader reasoning process. The paper further considers practical reforms designed to improve the communication and interpretation of quantitative evidence by experts, lawyers, judges, and jurors. In doing so, it seeks to clarify the role of statistical thinking in legal decision-making and identify practical implications for the presentation and interpretation of quantitative evidence in forensic contexts.

2. The Persuasive Power of Numbers

Quantitative evidence carries a distinctive authority within legal settings. Numbers appear precise, neutral, and resistant to bias. When presented in court, they often convey a sense of objectivity that qualitative evidence may lack [10,14]. This perception can be persuasive, particularly in adversarial contexts where clarity and certainty are highly valued. However, the persuasive force of numerical evidence does not necessarily reflect its interpretive reliability, nor the validity of the assumptions on which it is based. The problem is not the presence of numbers in the courtroom, but the misplaced belief that numerical expression equates to epistemic certainty.

This is the first dimension of the Statistical Theatre Model: the transformation of a numerical statement into a persuasive performance. Once expressed as a percentage, ratio, frequency, or probability, evidence may appear more settled than it actually is. This observation is consistent with longstanding scholarship examining the authority conferred upon quantitative evidence and expert testimony in legal settings [10,14,15]. The courtroom risk is therefore not simply that a number will be misunderstood, but that its numerical form will be treated as a proxy for reliability, neutrality, or certainty.

Part of this difficulty arises from the way in which statistical information is communicated. Numerical expressions are often presented in isolation, without sufficient explanation of the relationships they represent [1]. For example, a probability may be stated without reference to the conditions under which it was derived, or without distinguishing between different types of probability. Critically, distinctions such as the probability of the evidence given a hypothesis versus the probability that the hypothesis is true given the observed evidence are rarely made explicit in courtroom settings. Similarly, likelihood ratios may be presented without sufficient explanation of their relationship to competing propositions, creating a risk that they will be interpreted as direct statements about guilt or innocence rather than measures of evidential strength [1,12]. In such cases, the number itself becomes the focus, rather than the reasoning process it is intended to support.

This tendency is reinforced by the structure of adversarial advocacy. Courtroom argument is directed towards persuasion, and numerical evidence can operate as a powerful rhetorical device. A carefully selected statistic may appear to simplify complexity, but that simplification can obscure the assumptions, exclusions, and alternative explanations on which the number depends. The problem is not simplification itself, but simplification that removes the inferential scaffolding needed to interpret the evidence accurately.

The trial of O. J. Simpson provides a widely cited example of how statistical evidence can operate as persuasive theatre rather than structured reasoning. Complex DNA evidence, including population frequency estimates and match probabilities, was presented to the jury in ways that emphasised numerical magnitude while obscuring the assumptions and conditional relationships underlying those figures [16]. The evidential weight of the statistics was therefore inseparable from how they were framed and interpreted. The result was not simply disagreement about the evidence but competing interpretations of what the numbers meant. In this sense, the statistical evidence did not resolve uncertainty; it became part of the narrative contest itself. This dynamic—where numerical evidence contributes to competing interpretations rather than resolving them—continues to be observed in contemporary forensic contexts [5].

The persuasive effect of numbers is further amplified by cognitive factors. Individuals tend to associate numerical precision with accuracy, even when the underlying reasoning is incomplete. This reflects well-documented heuristics in judgment under uncertainty, whereby precision is conflated with validity [14,17]. More recent work suggests that the persuasive influence of numerical information may persist even when decision-makers are aware of potential sources of error, highlighting the resilience of quantitative framing effects in legal and forensic contexts [12]. This can lead to overconfidence in statistical evidence and reduced scrutiny of its assumptions. In legal contexts, where decisions carry significant consequences, such effects are particularly concerning.

This does not mean that numerical evidence is inherently unreliable. On the contrary, statistical reasoning can provide a powerful framework for evaluating uncertainty and weighing competing explanations. The difficulty arises when numbers are treated as self-explanatory. Without attention to the conditions under which a number was produced, the assumptions on which it rests, and the alternatives it does or does not exclude, numerical evidence may obscure rather than illuminate the issues before the court [9].

The danger, then, is not numerical evidence itself, but numerical evidence detached from its inferential structure. When that structure is not visible, the number may become theatrical: precise in appearance, persuasive in effect, but incomplete as reasoning. This is the central movement examined in the sections that follow.

3. Conditional Probability and the Prosecutor's Fallacy

A central difficulty in the interpretation of statistical evidence arises from the distinction between different forms of conditional probability [9,11,16]. In legal contexts, this distinction is frequently overlooked, giving rise to a class of errors that have been widely documented yet persist in practice. The most well-known of these is the prosecutor's fallacy. Despite decades of judicial guidance and academic critique, this error continues to appear in forensic and medico-legal contexts, indicating a persistent gap between statistical theory and legal application [6,11].

At its core, the problem involves confusion between two fundamentally different probabilities: the probability of observing the evidence given that a hypothesis is true, and the probability that the hypothesis is true given the observed evidence. These are not equivalent and cannot be substituted for one another without additional information [8]. They answer different inferential questions, and conflating them alters the direction of reasoning itself. This asymmetry lies at the heart of probabilistic reasoning and is frequently obscured in legal interpretation.

In courtroom settings, statistical evidence is often presented in terms of the probability of observing the evidence under a particular assumption. For example, an expert might testify that the probability of observing a particular DNA profile, assuming the defendant is not the source, is extremely low. This is a statement about the likelihood of the evidence under a specific condition. However, it is frequently interpreted as a statement about the likelihood that the defendant is guilty, given the evidence. This shift—from evidence given hypothesis to hypothesis given evidence—constitutes the prosecutor's fallacy [9]. In effect, the direction of inference is reversed without justification.

The distinction is not merely technical. It reflects a deeper issue about the structure of probabilistic reasoning. To move from the probability of the evidence under one proposition to an assessment of the relative plausibility of competing propositions, one must take into account additional factors, including prior information and the behaviour of the evidence under alternative explanations. This process is formalised in Bayesian reasoning, which provides a framework for updating beliefs in light of new evidence [1,8,12]. In many forensic disciplines, experts contribute to this process by evaluating the evidential strength of observations using likelihood ratios rather than expressing opinions on posterior probabilities or ultimate legal conclusions. In simplified terms, the question is not merely "how rare is this evidence?" but "how much more likely is this evidence if one explanation is true rather than another?". Without such a framework, any attempt to infer guilt directly from the rarity of the evidence is incomplete and, in practical terms, logically unsound. The problem is not mathematical complexity, but the omission of the conditions that give the mathematics meaning.

The persistence of the prosecutor's fallacy can be attributed in part to the intuitive appeal of rarity. A very small probability may appear compelling, particularly when expressed in numerical terms. However, rarity alone does not determine probative value. An event may be unlikely under one hypothesis, but equally or more unlikely under others. The evidential significance of a probability, therefore, depends upon comparative evaluation rather than numerical magnitude alone. Without considering these alternatives, the evidential significance of the probability cannot be properly assessed [10]. Rarity, in isolation, is therefore not evidence of guilt but a feature of the evidential distribution. This is why likelihood-based reasoning is important: it requires evidence to be evaluated comparatively, rather than treated as persuasive simply because it appears rare.

This issue is further complicated by the way probabilities are communicated. Numerical expressions are often presented without sufficient context, encouraging their interpretation as standalone indicators [14]. In such circumstances, the distinction between different conditional probabilities may not be apparent, even to those with legal training. The result is a form of reasoning in which the apparent clarity of the number obscures the structure of the argument, and the inferential pathway between evidence and conclusion is effectively bypassed. Importantly, research suggests that communication difficulties persist even where experts employ technically appropriate probabilistic frameworks, indicating that correct methodology alone does not guarantee correct interpretation. This difficulty is compounded by empirical findings that even trained forensic experts report challenges in communicating probabilistic reasoning clearly in courtroom settings [5].

The consequences of this confusion are well established. Misinterpretation of conditional probability has contributed to wrongful convictions and flawed expert testimony across multiple jurisdictions [8,16]. The persistence of the fallacy suggests that the difficulty is not simply a lack of statistical knowledge, but a failure to translate statistical reasoning into forms that legal decision-makers can use. In this sense, the prosecutor's fallacy is not only an error of probability; it is also a failure of evidential communication.

Addressing the prosecutor's fallacy requires more than the correction of individual statements. It demands a clearer understanding of the relationships between probabilities and the conditions under which they are defined. Only by engaging with this structure can numerical evidence be interpreted in a way that reflects its true evidential value, rather than as a substitute for reasoning itself. In practical terms, experts and courts must resist presenting a statistic as though it answers the ultimate legal question. The more appropriate task is to explain what the statistic compares, what assumptions it depends on, and what it does not prove.

4. Case Study: Sally Clark

The consequences of misinterpreting statistical evidence are illustrated starkly by the case of Sally Clark in the United Kingdom. Clark was convicted in 1999 of murdering her two infant sons, who had died suddenly at different times during infancy. Central to the prosecution's case was expert statistical testimony regarding the likelihood of two sudden infant deaths occurring within the same family [18–20].

The paediatrician Professor Roy Meadow testified that the probability of two cases of sudden infant death syndrome (SIDS) in an affluent, non-smoking family was approximately 1 in 73 million. This figure was derived by squaring the estimated probability of a single SIDS event, assuming the two deaths were independent. Meadow did explain the arithmetic of the calculation—stating that the probability of a single cot death in such a household was approximately 1 in 8500 and then squaring that figure—and the problem was therefore not that the number appeared without explanation, but that the independence assumption underpinning it was unsound. Presented in isolation, the number appeared extraordinarily small and was interpreted as indicating that the likelihood of the deaths being natural was negligible [21].

This reasoning was fundamentally flawed in several respects. First, and most importantly, the calculation assumed statistical independence between the two deaths, an assumption that lacked adequate empirical support. Subsequent analysis suggested that genetic, environmental, or other familial factors could increase the likelihood of recurrence, meaning the events were not statistically independent [20]. By treating them as such, the calculation artificially reduced the probability and exaggerated the apparent rarity of the outcome, thereby distorting the evidential baseline against which the deaths were assessed. The central error was therefore not simply one of arithmetic, but a misapplication of independence in a context where dependence was plausible [1].

Secondly, the inferential significance of the figure was liable to be misunderstood. The probability presented to the court was the probability of observing two sudden infant deaths under a particular explanatory framework, not the probability that any competing explanation was true. The distinction is central. A low probability of an event under one hypothesis does not, on its own, establish the probability of an alternative hypothesis. To make that inference would require consideration of prior probabilities and the likelihood of the evidence under competing explanations. More fundamentally, the evidential significance of two infant deaths could not be determined solely by examining the probability of two natural deaths. Proper probabilistic evaluation required comparison with alternative explanations, including the probability that two homicides would occur within the same household. Without such comparative assessment, the numerical estimate remained incomplete as a measure of evidential weight. While the case is most accurately understood as a failure of the independence assumption, it also intersected with the broader inferential error discussed earlier: the slide from a probability attached to evidence into an apparent probability of guilt [9,22].

The presentation of the statistics further compounded the problem. Although the arithmetic was straightforward, the inferential structure was not. The court was not provided with an adequate framework for evaluating whether the underlying assumptions were justified or how competing explanations should be weighed. Nor was the jury provided with a structured comparison of the relative explanatory power of competing propositions. As a result, the numerical estimate acquired a persuasive force that exceeded its evidential value.

Clark's conviction was eventually quashed in 2003 after it emerged that key medical evidence had not been disclosed and that the statistical testimony was unsound [18]. The case has since become a widely cited example of the dangers associated with the misuse of probability in legal settings. It highlights how numerical evidence, when misunderstood, can distort reasoning and contribute to serious miscarriages of justice.

The case also prompted renewed scrutiny of statistical reasoning in criminal proceedings and remains one of the most frequently cited examples of probabilistic error in legal decision-making.

The significance of the case lies not only in the specific errors made but also in what it reveals about the interaction between statistical reasoning and legal decision-making. The court was presented with a precise numerical estimate, but with a flawed conceptual framework for interpreting it. In the absence of sound probabilistic reasoning, the number was treated as a conclusion rather than as part of a conditional argument—a substitution of numerical assertion for structured inference [14]. The difficulty was therefore not merely that the number was wrong, but that the inferential task itself was framed incorrectly.

Viewed through the lens of Statistical Theatre, the Sally Clark case demonstrates how numerical precision can acquire persuasive authority independent of inferential strength. The figure of 1 in 73 million became rhetorically powerful not because it resolved the evidential questions before the court, but because it appeared to do so. The persuasive impact of the statistic ultimately exceeded the reasoning that supported it. The case, therefore, illustrates how numerical rarity can be mistaken for evidential conclusiveness when competing explanations are not explicitly evaluated.

A similar dynamic can be observed in other cases involving complex forensic evidence, where disagreement centres less on the data itself than on what the data permit one to conclude. Related concerns arose in the cases of Donna Anthony and Angela Cannings, both of whom were also convicted in the UK in circumstances shaped by Meadow's reasoning, before their convictions were later quashed. These cases reinforced the point that statistical error in court is rarely confined to a single number or a single trial; once flawed reasoning is accepted, it can travel [23,24].

The Sally Clark case demonstrates that the impact of statistical error is not confined to abstract reasoning. When misinterpreted, quantitative evidence can shape narratives of guilt and innocence in ways that are difficult to challenge once established. It underscores the need for careful articulation of assumptions, clear distinction between different types of probability, and a more structured approach to the use of statistical evidence in the courtroom, particularly where small probabilities carry disproportionate persuasive weight.

4.1. People v Collins and "Trial by Mathematics"

A related illustration appears in *People v Collins* [25], where the California Supreme Court rejected the prosecution's use of probability theory to identify a criminal defendant. There, the prosecution combined a series of unsupported probability estimates relating to characteristics such as race, facial hair, hair colour, and vehicle type, and invited the jury to treat the resulting figure as evidence of guilt. The court held that this approach was fundamentally flawed, both because the individual estimates lacked evidentiary foundation and because the traits were treated as independent when they were not [16,25].

Collins is important because it demonstrates that statistical error may arise long before a calculation is performed. The problem was not merely that probabilities were misunderstood, but that unsupported probabilities were created in the first place. Once assigned numerical values, those assumptions acquired an appearance of scientific legitimacy despite lacking an empirical foundation. Unlike the Sally Clark case, where the principal difficulty concerned the interpretation of probability and the validity of the independence assumption, Collins involved the assignment of numerical values that lacked adequate empirical support from the outset. The case, therefore, illustrates a distinct category of statistical error: not the misinterpretation of valid probabilities, but the construction of invalid probabilities.

The California Supreme Court's warning remains striking: mathematics may assist the trier of fact, but it must not "cast a spell" over the tribunal.

Viewed through the framework proposed in this paper, Collins represents an early example of Statistical Theatre. Numerical estimates were assigned, combined, and presented in a manner that created an appearance of analytical rigour while obscuring the weakness of the underlying assumptions. The result was not stronger evidence, but a stronger appearance of evidential strength. The case, therefore, serves as a reminder that statistical reasoning is only as reliable as the assumptions on which it is built [1]. More broadly, Collins demonstrates that quantitative evidence may become persuasive not only when probabilities are misunderstood, but also when numerical values are assigned an unwarranted degree of scientific credibility.

4.2. Contemporary Examples of Statistical Theatre

The issues illustrated by Sally Clark and *People v Collins* are often treated as historical lessons. However, the underlying problems remain highly relevant in contemporary legal and forensic practice. Advances in computing, artificial intelligence, forensic genomics, and algorithmic decision-making have increased the sophistication of quantitative evidence, but they have not eliminated the interpretive challenges associated with its use. In some respects, increasing complexity may amplify the risk that numerical outputs are accepted without adequate scrutiny. Recent scholarship suggests that as evidential systems become more computationally sophisticated, the challenge increasingly shifts from generating quantitative outputs to ensuring that those outputs can be meaningfully interpreted and evaluated by legal decision-makers [3,5,12].

One contemporary example involves algorithmic risk assessment tools used within criminal justice systems. Instruments such as the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) system generate estimates of recidivism risk intended to assist decision-making regarding sentencing, parole, and supervision. Although such systems produce numerical outputs that appear objective, critics have noted that the reliability and fairness of these estimates depend heavily on the assumptions, variables, and training data used to construct them [4,26]. The apparent precision of a risk score may therefore obscure substantial uncertainty regarding how that score was generated and what it actually represents. The issue is not that algorithmic tools are necessarily unreliable, but that the evidential significance of their outputs depends upon factors that may not be immediately visible to end users.

A related issue arises in forensic genomics and probabilistic genotyping. Modern software systems are capable of analysing complex DNA mixtures that would previously have been difficult or impossible to interpret. These systems often generate likelihood ratios expressing the relative strength of competing hypotheses regarding the source of a DNA sample [27,28]. Importantly, likelihood ratios are intended to assist evaluators by comparing the probability of the evidence under competing propositions rather than by expressing the probability that a particular proposition is true. While such approaches may represent a substantial methodological advance, their outputs can be difficult for judges, lawyers, and jurors to evaluate. The challenge is no longer simply understanding a probability, but understanding a probability produced through a highly complex computational process. As the underlying methodology becomes less transparent to non-specialists, the risk increases that confidence in the output may exceed understanding of the reasoning that produced it.

Artificial intelligence presents similar challenges. Machine-learning systems increasingly generate predictions, classifications, and recommendations across a range of legal and forensic contexts. Yet such systems frequently operate as "black boxes", producing outputs that may be difficult to explain in human terms [3,4]. Where decision-makers cannot readily examine the inferential pathway connecting inputs to outputs, there is a danger that numerical results acquire authority primarily because they are generated by sophisticated technology. In such circumstances, technological complexity may function in much the same way as statistical complexity, creating an appearance of objectivity that is not always matched by interpretive transparency. The resulting difficulty is not merely computational opacity, but the challenge of determining whether the reasoning embedded within the system can be adequately scrutinised, tested, and challenged.

Recent critiques of forensic science have raised related concerns regarding error rates and validation studies. For example, Cuellar et al. [29] argue that black-box studies used to evaluate forensic firearm comparison methods contain significant methodological weaknesses. Their analysis highlights an important point: even perfectly interpreted statistics cannot compensate for flawed underlying data or questionable assumptions. Statistical reasoning may clarify uncertainty, but it cannot rescue invalid measurements or poorly designed studies. In this respect, statistical interpretation and methodological validity remain complementary rather than interchangeable safeguards.

These contemporary examples demonstrate that Statistical Theatre is not confined to historical miscarriages of justice. The central issue remains the same: quantitative evidence may acquire persuasive force that exceeds its inferential value when the assumptions, limitations, and uncertainties underpinning it are insufficiently examined. Whether the source is a probability calculation, a risk assessment algorithm, a DNA interpretation system, or an artificial intelligence model, the challenge is ultimately one of reasoning rather than computation. The critical question is not whether a number has been produced, but whether the reasoning that produced it can be understood, evaluated, and challenged.

5. Witness Evidence and Bayesian Reasoning

The evaluation of witness evidence presents a distinct but related challenge in the interpretation of probabilistic information. Unlike forensic evidence, which is often accompanied by numerical estimates, witness testimony is typically assessed qualitatively. Nevertheless, its probative value can be understood in probabilistic terms [8,28,30]. Witness testimony occupies a different evidential position from forensic and statistical evidence. Whereas forensic experts often present information intended to assist the evaluation of competing propositions, witnesses typically provide factual observations or recollections. The present discussion does not suggest that witness testimony constitutes Statistical Theatre in the same manner as quantitative evidence. Rather, it is included because it illustrates a broader inferential principle: evidential weight depends upon the relationship between observations and competing explanations. The reliability of a witness does not establish the truth of a claim; rather, it affects the likelihood that the claim is true given the testimony. This requires distinguishing between the probability of observing the testimony if a hypothesis were true and the probability that the hypothesis is true given that testimony. This distinction is central to a coherent assessment of evidential weight.

Consider a witness who identifies a suspect as the perpetrator of an offence. The intuitive response may be to treat the identification as strong evidence of guilt, particularly if the witness appears confident and credible. However, the evidential value of the identification depends not only on the reliability of the witness, but also on how likely such an identification would be under alternative explanations. In other words, the key question is not simply whether the witness is accurate, but whether the identification meaningfully distinguishes between competing hypotheses [8].

This relationship can be expressed using Bayesian reasoning, which provides a framework for updating the probability of a hypothesis in light of new evidence [1,8,12]. In its simplest form, the probability of a hypothesis after considering evidence depends on two components: the prior likelihood of the hypothesis and the likelihood of the observed evidence under that hypothesis relative to alternatives. The likelihood ratio—comparing the probability of the evidence under competing explanations—captures the evidential contribution of the observation and provides a structured measure of how strongly the evidence discriminates between those explanations. It is this comparative structure, rather than the apparent reliability of the witness alone, that determines evidential weight.

Applied to witness testimony, this means asking two related questions: How likely is the witness to identify the suspect if the suspect is guilty, and how likely is the witness to identify the suspect if the suspect is innocent? The evidential value of the testimony depends on the difference between those two probabilities.

A highly reliable witness may still produce misleading evidence if the probability of false identification is non-negligible, particularly in situations where the prior probability of guilt is low [31]. In such cases, even accurate testimony may have limited probative value if it does not meaningfully distinguish between competing hypotheses. Conversely, even moderately reliable testimony may be probatively valuable if it significantly differentiates between competing explanations [8].

Difficulties arise when these relationships are not explicitly considered. Witness evidence is often evaluated in isolation, with emphasis placed on perceived credibility rather than on the structure of the inference. Confidence, demeanour, and consistency may influence judgments, but they do not directly quantify the likelihood of competing hypotheses [30,32]. Without a framework for integrating these factors, the assessment of testimony may rely on intuition rather than structured reasoning, and the inferential significance of the evidence may be misjudged.

This difficulty is compounded by cognitive constraints. Complex probabilistic reasoning places demands on attention and working memory that are not easily sustained in courtroom environments, increasing reliance on intuition and simplifying heuristics [33]. More recent research suggests that such inferential difficulties persist even where decision-makers are aware of common reasoning errors, indicating that the challenge is not merely educational but cognitive in nature [12].

The absence of numerical expression compounds the problem. While the probabilistic nature of forensic evidence is often made explicit, the probabilistic implications of witness testimony remain implicit. This can create an imbalance in how different forms of evidence are weighed. Numerical evidence may be overvalued due to its apparent precision, while qualitative evidence may be overinterpreted based on subjective impressions. In both instances, the underlying issue is a failure to engage with the conditional structure of the evidence [1].

The consequence is that confidence may be mistaken for evidential strength, despite research demonstrating that witness confidence and witness accuracy are not always equivalent [31]. Importantly, this differs from the persuasive influence of numerical evidence. In witness testimony, the concern is that confidence may be mistaken for accuracy; in quantitative evidence, the concern is that numerical precision may be mistaken for evidential strength. Although distinct, both errors arise from reliance upon surface features rather than inferential structure.

This distinction nevertheless helps explain why Statistical Theatre can be effective. The broader lesson is that decision-makers may place weight on features that appear persuasive—whether numerical precision, technological sophistication, or witness confidence—without fully evaluating the inferential relationships that give those features evidential meaning.

Bayesian reasoning does not require the routine use of formal calculations in the courtroom. Rather, it provides a conceptual framework for thinking about evidence in terms of conditional relationships. It emphasises that evidence should be evaluated according to how it alters the probability of competing hypotheses, not simply in terms of its standalone plausibility. This shift—from asking whether evidence appears credible to asking how it changes the balance between competing explanations—aligns closely with principled legal reasoning, even if it is not always explicitly articulated in practice [1].

The evaluation of witness evidence illustrates how probabilistic reasoning extends beyond numerical data. Whether expressed quantitatively or qualitatively, evidence derives its meaning from the way it relates to competing explanations. When these relationships are not explicitly recognised, there is a risk that evidential weight will be assigned on the basis of intuition or appearance rather than structured analysis.

Bayesian reasoning offers a means of addressing this issue by making those relationships explicit, thereby improving the coherence and transparency of legal decision-making and reducing the risk that persuasive features of testimony are mistaken for evidential strength.

The objective is not to transform judges, jurors, or lawyers into statisticians. Rather, it is to ensure that the reasoning underpinning evidential evaluation remains visible, transparent, and open to scrutiny.

6. Base Rates and the Denominator Problem

A further source of error in the interpretation of statistical evidence arises from neglecting base rates, sometimes referred to as the denominator problem [34]. Base rates represent the underlying frequency with which an event or characteristic occurs within a relevant population and therefore provide essential context for interpreting probabilistic statements. When base rates are ignored, the relationship between the evidence and the hypothesis is mischaracterised, and the evidential significance of a probability may be substantially distorted. In such circumstances, attention is directed to the numerator—the apparent rarity of the event—while the denominator, which gives that rarity meaning, is effectively disregarded.

The importance of base rates can be illustrated by considering identification evidence. Suppose a witness identifies an individual from a large population. Even if the identification procedure is relatively accurate, the probability that the identified individual is in fact the correct person depends critically on how many individuals in the population could plausibly match the description. If the population is large, the number of potential matches increases, and the likelihood of a coincidental or erroneous identification correspondingly rises [31]. In such cases, the evidential value of the identification depends not only on accuracy but on the ratio of true identifications to false positives across the population. Without reference to this broader evidential context, the identification may appear more probative than it is.

In Bayesian terms, base rates form part of the prior probability of a hypothesis. This means that the significance of evidence depends not only on the evidence itself, but also on how plausible the explanation was before the evidence was observed [1,34]. Importantly, the role of a prior probability is not to determine the outcome of an inference but to provide the context in which new evidence can be evaluated. When base rates are ignored, probabilities may appear more compelling than they actually are, resulting in a distorted assessment of evidential strength.

The denominator problem becomes particularly pronounced in cases involving rare events. When an event is described as highly unlikely, attention is often focused on the small probability itself, rather than on the number of opportunities for that event to occur. For example, an event with a probability of 1 in 1000 may appear rare in isolation, but if there are many thousands of comparable opportunities for the event to arise, its occurrence becomes less surprising. Rarity must therefore be interpreted relative to the size of the relevant population or number of trials. Without this reference point, rarity can appear more informative than it actually is. This illustrates a broader principle: probabilities derive evidential meaning from context rather than magnitude alone.

In legal contexts, this issue frequently arises in the presentation of forensic evidence. A match probability may be expressed as a very small number, suggesting that the evidence strongly implicates a particular individual. However, if the relevant population is large, the expected number of individuals who would share the same characteristic may not be negligible [16]. In such circumstances, the probability of a match does not translate directly into the probability of guilt. The failure to consider this denominator effect can lead to overestimation of the evidential value of the match. This concern aligns with the reasoning criticised in *People v Collins* [25], where probabilistic claims were advanced without adequate consideration of the number of potential alternative matches, thereby inflating the apparent strength of the evidence. More broadly, the evidential significance of a match depends upon comparison with competing explanations rather than upon the rarity of the observed characteristic alone.

The neglect of base rates is not simply a mathematical oversight; it reflects a broader cognitive tendency to focus on salient information while neglecting contextual information [33]. Related research in judgment and decision-making has similarly demonstrated that individuals often privilege vivid or intuitively compelling information over background statistical information, even when the latter is highly relevant to the inferential task [13,14]. A striking numerical value may attract attention precisely because it appears informative, even when critical contextual information is missing.

Addressing this problem requires a shift in how probabilistic information is presented and interpreted. Rather than emphasising isolated probabilities, greater attention must be given to the conditions under which those probabilities are meaningful. This includes explicit consideration of the relevant population, the number of opportunities for the event to occur, and the prior plausibility of competing hypotheses, thereby restoring the link between numerical values and their inferential context.

Viewed through the framework of Statistical Theatre, base rate neglect demonstrates how numerical evidence can appear persuasive while remaining incomplete. The problem is not that the number is wrong, but that the context required to interpret it has been removed. A probability, however precise, cannot be properly understood without reference to the population, assumptions, and competing explanations from which it derives. Recognising the role of base rates is therefore essential to ensuring that quantitative evidence reflects its true evidential significance rather than merely its persuasive impact.

7. DNA Evidence and Misinterpretation

DNA evidence is often regarded as one of the most reliable forms of forensic proof. Advances in genetic profiling have enabled highly discriminating comparisons between biological samples, and statistical estimates are routinely used to quantify the strength of a match [27,28]. These estimates are typically expressed as match probabilities or likelihood ratios, conveying the probability of observing the genetic profile under specified assumptions. That is, they describe the behaviour of the evidence under particular hypotheses rather than the probability of those hypotheses themselves. Despite their scientific foundation, such measures are frequently misunderstood in legal contexts.

A common point of confusion arises from the interpretation of match probabilities. When a DNA expert reports that the probability of a coincidental match is extremely small, this is a statement about the likelihood of the evidence under a particular hypothesis—usually that the DNA originated from someone other than the suspect. In practical terms, this reflects how likely the observed DNA profile would be if the suspect were not the source. However, this figure is often interpreted as indicating the probability that the suspect is not the source of the DNA, or conversely, that the suspect is almost certainly guilty. This involves an unjustified shift from the probability of the evidence under a hypothesis to the probability of the hypothesis given the evidence, mirroring the error discussed earlier in relation to conditional probability [9,11].

The distinction is critical. A small match probability does not, on its own, establish the probability that any particular proposition is true. Rather, the evidential significance of a DNA match depends on how strongly it distinguishes between competing explanations. Likelihood ratios provide a more appropriate framework because they explicitly compare the probability of the evidence under alternative hypotheses [1,12]. Importantly, likelihood ratios are intended to assist evaluators by expressing the relative support that the evidence provides for competing propositions, rather than by providing direct estimates of guilt, innocence, or posterior probabilities.

In this sense, the evidential strength of a DNA match lies in its capacity to discriminate between hypotheses, not in the magnitude of a single probability. Even so, the interpretation of likelihood ratios requires an understanding of what is being compared and under what conditions—a requirement that is not always satisfied in adversarial settings.

Further complications arise from the role of population statistics. DNA match probabilities are typically derived from databases that estimate the frequency of particular genetic profiles within a population [27].

These estimates depend on assumptions about population structure, independence of genetic markers, and sampling methods. While such assumptions are often reasonable, they are not always made explicit in courtroom presentations. Consequently, the reported probability may be treated as unconditional rather than contingent on these assumptions. What appears to be a single definitive figure is therefore the product of multiple underlying assumptions, each of which shapes its interpretation [1].

The denominator problem discussed in the previous section is also relevant in this context. Even very small match probabilities may correspond to a non-negligible number of individuals within a large population who share similar genetic characteristics. Without considering the size and composition of the relevant population, the evidential weight of a DNA match may be overstated [16]. In such cases, the rarity of the profile does not uniquely identify the suspect but reflects its distribution within the population. This is particularly important in cases where the suspect was identified through database searches or where the pool of potential contributors is large. Database trawling cases provide a useful illustration of this problem because the probability of identifying a coincidental match may increase as the size of the searched database increases [35].

The persuasive impact of DNA evidence is considerable. Its scientific basis, technological sophistication, and numerical expression can create an impression of near certainty, particularly when probabilities are expressed in extremely small terms [14]. Yet this apparent precision may obscure the conditional nature of the evidence and encourage overconfidence in the conclusion drawn from it.

A DNA match increases the likelihood of a particular hypothesis relative to alternatives, but it does not resolve the question of guilt in isolation. DNA evidence, like all evidence, must be interpreted within a broader evidential context, including issues of transfer, contamination, timing, and alternative explanations [28].

For this reason, DNA evidence alone is rarely sufficient to determine guilt. Its probative value depends on how it is integrated with other forms of evidence, including opportunity, motive, context, and competing explanations regarding how and when the DNA was deposited. Even highly discriminating genetic evidence cannot independently resolve those questions.

Modern probabilistic genotyping systems further illustrate this challenge. Contemporary software can analyze complex DNA mixtures and generate likelihood ratios that would have been difficult to calculate manually [27,28]. While these systems represent important scientific advances, they also increase the distance between the numerical output and the decision-maker's understanding of how that output was generated. As forensic interpretation becomes increasingly computational, the risk grows that confidence in the result may exceed understanding of the reasoning process itself. Importantly, this concern does not arise because probabilistic genotyping is inherently unreliable, but because increasing computational sophistication may create additional challenges for transparency, communication, and evaluation by non-specialists.

Viewed through the framework of Statistical Theatre, DNA evidence demonstrates how highly reliable scientific methods can nevertheless be vulnerable to interpretive error. The issue is not the validity of DNA profiling itself, but the tendency to treat statistical outputs as conclusions rather than as components of a broader inferential process. The probative value of DNA evidence lies not in the number itself, but in the reasoning framework within which it is interpreted. Ensuring that DNA evidence is presented and evaluated within that framework is essential to maintaining the integrity of legal decision-making.

8. Cognitive Constraints and Statistical Reasoning

The persistence of errors in the interpretation of statistical evidence cannot be explained solely by a lack of technical knowledge. Even where numerical information is presented clearly, misunderstandings frequently arise [33]. This suggests that the difficulty lies not only in the mathematics itself, but in the cognitive processes used to interpret it within legal contexts. The challenge is therefore not merely informational, but cognitive and structural. Research in legal psychology, decision-making, and forensic science increasingly suggests that difficulties with probabilistic reasoning may persist even among educated and experienced decision-makers, highlighting the importance of how statistical information is communicated and contextualised [12,14].

Research in cognitive psychology suggests that people frequently rely on two broad modes of thinking. One mode is relatively fast, intuitive, and automatic, while the other is slower, more analytical, and effortful [33]. Probabilistic reasoning often requires the latter. Yet courtroom environments frequently encourage rapid judgments, simplified narratives, and reliance on intuitive impressions, creating conditions in which statistical information may be misunderstood despite being technically correct.

In the context of legal decision-making, statistical evidence is often approached procedurally. Numerical outputs—such as probabilities, percentages, or likelihood ratios—are treated as results to be accepted or rejected, rather than as components of a broader inferential process [8]. In effect, the number becomes the endpoint of reasoning rather than the starting point for analysis. This can lead to a form of reasoning in which the presence of a number substitutes for analysis—a shift from interpretation to acceptance. What is lost in this transition is the conditional structure that gives the number its meaning. The structure of the argument, including its assumptions and conditional dependencies, may remain implicit or unexamined.

This tendency is reinforced by the cognitive demands associated with probabilistic reasoning. Evaluating conditional probabilities requires decision-makers to consider prior probabilities, alternative explanations, and competing interpretations simultaneously. Such tasks place substantial demands on working memory and analytical reasoning. Under these conditions, reliance on cognitive shortcuts, or heuristics, becomes both predictable and difficult to avoid [17]. While heuristics are often useful, they may also produce systematic errors when applied to probabilistic information. Research on probabilistic reasoning has further demonstrated that errors such as base rate neglect, confusion of conditional probabilities, and misinterpretation of statistical significance are not isolated mistakes but recurring patterns of judgment under uncertainty [11,13].

The adversarial nature of legal proceedings may further exacerbate these difficulties. Evidence is often presented in a manner designed to persuade, with emphasis placed on clarity and impact. Numerical statements that appear precise and definitive may be particularly influential, even when their interpretation depends on underlying assumptions that are not fully articulated [14]. In such contexts, persuasive presentation may displace analytical scrutiny. The result is that evidential form may take precedence over evidential substance.

Importantly, these issues are not confined to jurors. Judges, lawyers, forensic practitioners, and expert witnesses may also encounter difficulties when engaging with probabilistic reasoning outside their areas of expertise [16]. Research within forensic science has similarly demonstrated the influence of contextual information, cognitive bias, and inferential shortcuts on professional decision-making, prompting the development of structured approaches designed to reduce such effects [36]. The development of Linear Sequential Unmasking–Expanded (LSU-E), for example, reflects growing recognition that even experienced forensic practitioners may be influenced by contextual information and cognitive biases that operate independently of technical competence [36].

This was illustrated in the Sally Clark case, where expert statistical reasoning was presented with apparent authority but rested on flawed assumptions, demonstrating that even highly credentialed individuals may misapply probabilistic concepts when the underlying structure is not fully interrogated. The lesson is not that expertise lacks value, but that expertise alone does not eliminate the need for transparent reasoning and critical evaluation of assumptions.

The increasing complexity of contemporary forensic evidence compounds these challenges. Decision-makers are now routinely asked to evaluate probabilistic genotyping systems, algorithmic risk assessments, and other computationally intensive forms of evidence. As complexity increases, there is a corresponding risk that trust in the output may exceed understanding of the process that produced it.

Viewed through the framework of Statistical Theatre, these cognitive constraints help explain why numerical evidence can acquire persuasive authority independent of its inferential value. A precise figure, likelihood ratio, or probability estimate may appear compelling because it reduces complexity and creates an impression of certainty. Yet the very features that make such information persuasive may also discourage deeper examination of the assumptions on which it depends. The difficulty is therefore not merely computational. It is simultaneously cognitive, structural, and institutional, arising from the interaction between human reasoning processes, legal procedures, and the presentation of quantitative evidence.

Addressing these cognitive limitations does not require transforming legal practitioners into mathematicians. Rather, it involves recognising the types of reasoning that statistical evidence demands and ensuring that its presentation supports, rather than hinders, conceptual understanding. By making the structure of probabilistic arguments more explicit and by aligning their presentation with known cognitive constraints, it may be possible to reduce the risk of misinterpretation and improve the quality of decision-making in legal contexts. The objective is not to increase complexity, but to render complexity intelligible without sacrificing accuracy.

9. The Statistical Theatre Model

The preceding sections have examined a series of apparently distinct problems involving quantitative evidence. These have included the prosecutor's fallacy, base rate neglect, misunderstandings of DNA evidence, errors in witness evaluation, flawed assumptions regarding statistical independence, and the increasing complexity of algorithmic and computational forms of evidence. Although these problems arise in different contexts, they share a common feature: numerical information acquires persuasive authority that exceeds its inferential value.

This paper proposes the Statistical Theatre Model as a framework for understanding this phenomenon. The underlying issues examined throughout this paper are not new. Previous scholarship has explored probabilistic fallacies, evidential communication, cognitive bias, expert testimony, and the persuasive influence of numerical information in legal decision-making [10,12,14,15]. The contribution of the present model is therefore not the identification of a previously unrecognised problem, but the integration of these recurring issues within a single explanatory framework. Statistical Theatre refers to situations in which quantitative evidence becomes persuasive not primarily because of the strength of the reasoning it supports, but because of the appearance of objectivity, precision, and certainty conveyed by numerical expression. The issue is not that the mathematics is necessarily incorrect. Rather, the persuasive force of the number may become detached from the assumptions, conditions, and inferential processes that give the number its meaning.

The model does not suggest that statistical evidence is inherently misleading. On the contrary, statistical reasoning remains one of the most powerful tools available for evaluating uncertainty and comparing competing explanations [1,8]. The concern arises when numerical outputs are treated as conclusions rather than as components of a broader reasoning process. Under such circumstances, the authority associated with the number may exceed the evidential value it legitimately possesses.

The Statistical Theatre Model proposes that this process typically occurs through a series of stages.

9.1. Stage 1: Evidence Is Quantified

The first stage involves the conversion of observations into numerical form. This may occur through statistical analysis, probability estimates, likelihood ratios, risk scores, algorithmic classifications, or other quantitative techniques. Quantification is often beneficial because it imposes structure on uncertainty and enables systematic comparison between competing explanations. However, the act of quantification may also create an impression of increased objectivity, regardless of whether the underlying assumptions are fully understood.

9.2. Stage 2: Complexity Is Compressed

Once quantified, complex evidential relationships are frequently condensed into a numerical output. A DNA likelihood ratio, a recidivism score, a match probability, or a diagnostic estimate may summarise a large amount of information in a single figure. This compression is often necessary for practical communication. However, it also reduces visibility of the assumptions, limitations, and alternative explanations that contributed to the calculation. Complexity is not eliminated; it is hidden.

9.3. Stage 3: Persuasive Authority Emerges

The numerical output may then acquire persuasive authority because of its apparent precision. Research in cognitive psychology demonstrates that people often associate numerical specificity with accuracy and reliability, even when the underlying reasoning remains uncertain [14,33]. Numbers appear objective, neutral, and scientific. As a consequence, they may be granted a degree of credibility that exceeds what would be afforded to equivalent qualitative claims.

9.4. Stage 4: Inferential Shortcuts Occur.

At this stage, attention shifts from the reasoning process to the numerical result itself. Questions regarding assumptions, alternative explanations, prior probabilities, error rates, and contextual factors receive less scrutiny than the numerical output. Inferential shortcuts become more likely. The prosecutor's fallacy, base rate neglect, misunderstandings of likelihood ratios, and overreliance on algorithmic outputs can all be understood as manifestations of this stage [9,11].

9.5. Stage 5: Numerical Conclusions Replace Probabilistic Reasoning

The final stage occurs when a numerical statement is treated as though it resolves the evidential question. A probability becomes interpreted as proof, a likelihood ratio becomes interpreted as direct support for a legal conclusion, or a risk score becomes interpreted as certainty. At this point, the distinction between evidence and conclusion begins to collapse. The numerical output is no longer functioning as an aid to reasoning; it has become a substitute for reasoning.

The value of the Statistical Theatre Model lies in its ability to unify problems that are often treated separately. The Sally Clark case, *People v Collins*, DNA interpretation errors, witness-identification problems, algorithmic risk assessments, and contemporary concerns regarding artificial intelligence all involve different forms of evidence. Yet each illustrates the same underlying process: numerical information acquiring persuasive force independent of its actual inferential value.

Importantly, the model also helps explain why these errors persist despite repeated judicial warnings and academic commentary. The problem is not simply a lack of statistical education. Rather, it emerges from the interaction between cognitive limitations, adversarial advocacy, institutional pressures, and the persuasive properties of numerical communication itself. Even highly educated decision-makers may be vulnerable when complexity is compressed into a form that appears definitive while concealing uncertainty.

The model also has practical implications. By identifying the stages through which quantitative information may acquire unwarranted persuasive authority, it provides a framework for evaluating how statistical evidence is communicated, challenged, and interpreted within legal proceedings. The model therefore operates not only as a descriptive account of recurrent evidential problems, but also as a heuristic for identifying points at which interpretive safeguards may be introduced.

The Statistical Theatre Model therefore shifts attention away from individual statistical mistakes and towards the broader conditions that make such mistakes possible. In doing so, it provides a framework for understanding not only how quantitative evidence can mislead, but also how legal systems might better communicate, evaluate, and regulate probabilistic information. The model does not argue against the use of statistics in the courtroom. Instead, it argues for a more transparent engagement with the assumptions, limitations, and inferential structures that statistical evidence necessarily entails.

10. Practical Reforms for Courts and Experts

The issues identified in this paper have practical implications for the way quantitative evidence is presented and evaluated in legal proceedings. While statistical reasoning provides a powerful framework for analysing uncertainty, its value depends on how effectively it is communicated and understood [1,8].

Improving this process does not require greater mathematical sophistication. Rather, it requires greater transparency regarding the assumptions, limitations, and inferential structures that underpin quantitative evidence. As the Statistical Theatre Model suggests, many interpretive failures arise not because the mathematics is incorrect, but because the reasoning supporting the mathematics is insufficiently visible to those tasked with evaluating it. The practical value of the model lies in its ability to identify points at which quantitative information may become detached from its inferential foundations, thereby highlighting opportunities for intervention and reform.

A central requirement is the explicit articulation of assumptions. Probabilities are always conditional on specific premises, whether these relate to population characteristics, independence of events, or alternative explanations. When such assumptions remain implicit, numerical statements may be interpreted as absolute measures rather than as context-dependent estimates. Experts therefore have a responsibility to make clear not only the numerical results they present, but the conditions under which those results hold—including the extent to which those conditions may reasonably be contested or uncertain. Without such articulation, the apparent precision of a number may conceal the fragility of the reasoning on which it depends.

Equally important is the manner in which probabilities are communicated. As discussed throughout this paper, many interpretive errors arise from the tendency to treat statistical outputs as though they directly answer legal questions. Expert evidence should instead emphasise comparison rather than conclusion. Rather than presenting a probability as evidence of guilt or innocence, experts should explain how the observed evidence behaves under competing explanations. This approach aligns more closely with likelihood-based reasoning [1] and reduces the risk that statistical evidence will be mistaken for a determination of the ultimate issue before the court. Where likelihood ratios are used, their role as measures of evidential support, rather than as determinations of legal conclusions, should be made explicit.

The presentation of statistical evidence should also account for the role of base rates and relevant populations. Isolated probabilities, particularly when expressed as very small values, can be misleading if not situated within an appropriate context [14]. Providing information about the size and characteristics of the relevant population can assist decision-makers in understanding the broader significance of a probability. Where possible, frequencies or comparative scenarios may offer a more intuitive means of conveying this information without sacrificing accuracy [13]. Such representations align more closely with how individuals naturally process probabilistic information. The objective is not simplification, but faithful representation of the underlying relationships.

Judicial directions also play a critical role in shaping how quantitative evidence is understood. Courts routinely provide guidance on complex legal concepts, and there is no reason that common statistical misconceptions should be treated differently. Directions that explicitly address issues such as the prosecutor's fallacy, base rate neglect, and the distinction between probabilities and conclusions may assist decision-makers in maintaining an appropriate focus on the reasoning process rather than the numerical output alone [16]. The development of standardised judicial guidance, informed by both legal and statistical expertise, may therefore represent a practical safeguard against the misuse of apparent certainty.

There is also a role for restraint in the use of numerical evidence. Not all aspects of a case are amenable to precise quantification, and the introduction of numerical estimates may create a false impression of certainty. Where probabilities are presented, they should reflect meaningful distinctions between competing explanations, rather than being used as rhetorical devices—maintaining a clear distinction between evidential contribution and persuasive impact. The objective is not to maximise persuasive force, but to preserve the integrity of the reasoning process.

Finally, the increasing complexity of contemporary forensic evidence highlights the importance of interdisciplinary engagement. Legal decision-makers are increasingly required to evaluate evidence generated through advanced statistical methods, computational modelling, probabilistic genotyping, and algorithmic systems. Collaboration among legal scholars, statisticians, psychologists, and forensic scientists may help ensure that such evidence is both scientifically rigorous and meaningfully interpretable in legal settings [4,28]. The challenge is therefore not merely technical, but communicative and conceptual.

The practical challenge is not to eliminate uncertainty, but to represent it faithfully. Statistical evidence, when properly framed, can enhance the quality of legal decision-making. When misapplied, it risks distorting the inferential process by creating an appearance of certainty that exceeds what the evidence can justify. The distinction lies not in whether quantitative methods are employed, but in whether the assumptions, limitations, competing explanations, and reasoning structures underpinning those methods remain visible and open to scrutiny. Reducing Statistical Theatre, therefore, requires greater transparency, clearer communication, and a sustained commitment to evidential reasoning over numerical persuasion.

11. Conclusions

Quantitative evidence has become an increasingly prominent feature of modern legal proceedings, offering the promise of precision in the evaluation of uncertainty [16]. However, as this paper has demonstrated, the difficulties associated with such evidence lie not in the mathematics itself, but in its interpretation. Probabilities do not provide direct answers to questions of guilt or innocence; they describe relationships between evidence and competing explanations under specified conditions. They are, therefore, tools of inference, not conclusions in themselves. When these relationships are obscured or misunderstood, the resulting errors are not merely technical, but substantive, with potentially significant legal consequences.

Across the examples examined in this paper—including the Sally Clark case, *People v Collins*, DNA interpretation, witness evidence, base rate neglect, and contemporary algorithmic systems—a common pattern emerges. Quantitative evidence may acquire persuasive authority that exceeds its actual inferential value. The issue is not that numerical evidence is inherently unreliable, but that numerical expression can create an appearance of certainty that is not always warranted by the underlying reasoning.

The Statistical Theatre Model proposed in this paper provides a framework for understanding this process. Rather than viewing probabilistic errors as isolated mistakes, the model integrates a range of recognised evidential and interpretive problems within a broader explanatory framework in which quantification, compression of complexity, persuasive authority, and inferential shortcuts interact to produce overconfidence in numerical outputs. The contribution of the model is therefore not the identification of a new category of error, but the synthesis of recurring problems that have traditionally been examined in relative isolation. This framework helps explain why similar interpretive failures continue to arise across different forms of evidence despite decades of judicial guidance and academic commentary.

The implications for legal practice are equally clear. Improving the use of quantitative evidence does not require greater mathematical sophistication, but greater transparency regarding assumptions, clearer communication of uncertainty, and sustained attention to the inferential structures that numerical outputs represent. The objective is not to eliminate uncertainty, but to ensure that uncertainty is represented faithfully and interpreted appropriately.

Statistical reasoning has much to offer the legal system, particularly in its capacity to clarify uncertainty and evaluate competing explanations. Its effectiveness, however, depends on whether decision-makers remain focused on the reasoning that numerical evidence is intended to support. Where numbers become substitutes for that reasoning, quantitative evidence risks functioning less as a tool of clarification and more as a form of Statistical Theatre. The challenge for courts, experts, and legal practitioners is therefore not to produce more numbers, but to ensure that numerical evidence remains subordinate to the inferential process it is intended to inform—and never mistaken for it.

More broadly, the Statistical Theatre Model highlights the importance of preserving the distinction between evidential support and legal conclusion. Quantitative evidence can assist legal decision-making by clarifying uncertainty and discriminating between competing explanations, but only when its assumptions, limitations, and inferential foundations remain visible. The enduring lesson of the cases and examples examined throughout this paper is that numerical precision is most valuable when it supports reasoning rather than replacing it.

Acknowledgments

Rebecca C. K. Ward acknowledges James Cook University for its ongoing institutional support.

Author Contributions

J.S.C.: Conceptualization, Methodology, Investigation, Formal Analysis, Data Curation, Writing—Original Draft Preparation, Writing—Review & Editing, Validation, and Project Administration. R.C.K.W.: Conceptualization, Methodology, Investigation, Formal Analysis, Data Curation, Writing—Original Draft Preparation, Writing—Review & Editing, Validation, and Project Administration. Both authors contributed equally to the work and have read and agreed to the published version of the manuscript.

Ethics Statement

Not applicable. This study did not involve human participants, animals or identifiable personal data.

Informed Consent Statement

Not applicable. This study did not involve human participants.

Data Availability Statement

No new data were created or analysed for this conceptual article.

Funding

The research received no external funding, and the article processing charge was waived by the publisher.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Aitken CGG, Taroni F. *Statistics and the Evaluation of Evidence for Forensic Scientists*, 2nd ed.; Wiley: Chichester, UK, 2004. DOI:10.1002/0470011238
2. Robertson B, Vignaux GA. *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*; Wiley: Hoboken, NJ, USA, 1995.
3. Guadamuz A. The EU's Artificial Intelligence Act and copyright. *J. World Intellect. Prop.* **2025**, *28*, 213–219. DOI:10.1111/jwip.12330
4. Hildebrandt M. *Law for Computer Scientists and Other Folk*; Oxford University Press: Oxford, UK, 2018.

5. Dawson HR, LaBat D, Sparacino M, Marciano M, Schreiber Compo N. DNA analysts' experiences with human factors: A quantitative and qualitative snapshot. *J. Forensic Sci.* **2025**, *70*, 1414–1423. DOI:10.1111/1556-4029.70064
6. Wilson AL, Cumbers S. Statistics is important in medical misconduct cases. *BMJ* **2026**, *392*, s226. DOI:10.1136/bmj.s226
7. Devlin K. *Introduction to Mathematical Thinking*; Keith Devlin: Palo Alto, CA, USA, 2012.
8. Fenton N, Neil M. Avoiding probabilistic reasoning fallacies in legal practice using Bayesian networks. *Aust. J. Leg. Philos.* **2011**, *36*, 114–151. Available online: https://access.heinonline.com/HOL/LandingPage?handle=hein.journals/ajlph36&div=7&id=&page= (accessed on 29 May 2026).
9. Thompson WC, Schumann EL. Interpretation of statistical evidence in criminal trials: The prosecutor's fallacy and the defense attorney's fallacy. *Law Hum. Behav.* **1987**, *11*, 167–187. DOI:10.1007/BF01044641
10. Koehler JJ. On conveying the probative value of DNA evidence: Frequencies, likelihood ratios, and error rates. *Univ. Colo. Law Rev.* **1996**, *67*, 859–886. Available online: https://access.heinonline.com/HOL/LandingPage?handle=hein.journals/ucollr67&div=38&id=&page= (accessed on 23 July 2026).
11. Dahlman C. A systematic account of probabilistic fallacies in legal fact-finding. *Int. J. Evid. Proof* **2024**, *28*, 45–64. DOI:10.1177/13657127231209019
12. Lagnado DA. *Explaining the Evidence: How the Mind Investigates the World*; Cambridge University Press: Cambridge, UK, 2021.
13. Gigerenzer G, Hoffrage U. How to improve Bayesian reasoning without instruction: Frequency formats. *Psychol. Rev.* **1995**, *102*, 684–704. DOI:10.1037/0033-295X.102.4.684
14. Gigerenzer G. *Calculated Risks: How to Know When Numbers Deceive You*; Simon & Schuster: New York, NY, USA, 2002.
15. Aitken CGG. Lies, damned lies and expert witnesses. *Math. Today* **1996**, *32*, 76–80. Available online: https://cdn.ima.org.uk/wp/wp-content/uploads/2024/09/Lies-Damned-Lies-and-Expert-Witnesses_Mathematics-Today_1996.pdf (accessed on 29 May 2026).
16. Kaye DH, Bernstein DE, Mnookin JL. *The New Wigmore: A Treatise on Evidence: Expert Evidence*, 2nd ed.; Wolters Kluwer Law & Business/Aspen Publishers: Austin, TX, USA, 2011.
17. Tversky A, Kahneman D. Judgment under uncertainty: Heuristics and biases. *Science* **1974**, *185*, 1124–1131. DOI:10.1126/science.185.4157.1124
18. R v Clark [2003] EWCA Crim 1020. Available online: https://www.bailii.org/ew/cases/EWCA/Crim/2003/1020.html (accessed on 22 July 2026).
19. Donnelly P. Appealing statistics. *Significance* **2005**, *2*, 46–48. DOI:10.1111/j.1740-9713.2005.00089.x
20. Hill R. Reflections on the cot death cases. *Significance* **2005**, *2*, 13–16. DOI:10.1111/j.1740-9713.2005.00077.x
21. Goldacre B. *Bad Science*; Fourth Estate: London, UK, 2008.
22. Dawid AP. Probability and proof. In *Analysis of Evidence*, 2nd ed.; Anderson TJ, Schum DA, Twining W, Eds.; Cambridge University Press: Cambridge, UK, 2005; pp. 381–404.
23. R v Cannings [2004] EWCA Crim 1. Available online: https://www.bailii.org/ew/cases/EWCA/Crim/2004/1.html (accessed on 23 July 2026).
24. R v Anthony [2005] EWCA Crim 952. Available online: https://www.bailii.org/ew/cases/EWCA/Crim/2005/952.html (accessed on 23 July 2026).
25. People v Collins, 68 Cal. 2d 319 (1968). Available online: https://law.justia.com/cases/california/supreme-court/2d/68/319.html (accessed on 23 July 2026).
26. Angwin J, Larson J, Mattu S, Kirchner L. Machine bias. ProPublica, 23 May 2016. Available online: https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (accessed on 23 July 2026).
27. Butler JM. *Advanced Topics in Forensic DNA Typing: Interpretation*; Academic Press: San Diego, CA, USA, 2015. DOI:10.1016/C2011-0-07649-4
28. Robertson B, Vignaux GA, Berger CEH. *Interpreting Evidence: Evaluating Forensic Science in the Courtroom*, 2nd ed.; Wiley: Chichester, UK, 2016. DOI: 10.1002/9781118492475
29. Cuellar M, Vanderplas S, Luby A, Rosenblum M. Methodological problems in every black-box study of forensic firearm comparisons. *Law Probab. Risk* **2024**, *23*, 1–27. DOI:10.1093/lpr/mgae015
30. Croucher JS. Witness evidence: Can it be trusted? Med. Law **2005**, *24*, 233–242. Available online: https://access.heinonline.com/HOL/LandingPage?handle=hein.journals/mlv24&div=24&id=&page= (accessed on 29 May 2026).

31. Wells GL, Olson EA. Eyewitness testimony. *Annu. Rev. Psychol.* **2003**, *54*, 277–295. DOI:10.1146/annurev.psych.54.101601.145028
32. Loftus EF. Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learn. Mem.* **2005**, *12*, 361–366. DOI:10.1101/lm.94705
33. Kahneman D. *Thinking, Fast and Slow*; Farrar, Straus and Giroux: New York, NY, USA, 2011.
34. Koehler JJ. The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behav. Brain Sci.* **1996**, *19*, 1–17. DOI: 10.1017/S0140525X00041157
35. Balding DJ, Donnelly P. Evaluating DNA profile evidence when the suspect is identified through a database search. *J. Forensic Sci.* **1996**, *41*, 603–607. DOI:10.1520/JFS13961J
36. Dror IE, Kukucka J. Linear Sequential Unmasking–Expanded (LSU-E): A general approach for improving decision making as well as minimizing noise and bias. *Forensic Sci. Int. Synerg.* **2021**, *3*, 100161. DOI:10.1016/j.fsisyn.2021.100161