



A systematic analysis of misleading evidence in unsafe rulings in England and Wales

Nadine M. Smit^{a,*}, Ruth M. Morgan^a, David A. Lagnado^b

^a Department of Security and Crime Science, Centre for the Forensic Sciences, University College London, 35 Tavistock Square, London WC1H 9EZ, United Kingdom

^b Department of Experimental Psychology, University College London, 26 Bedford Way, London WC1H 0AP, United Kingdom

ARTICLE INFO

Keywords:

Unsafe rulings
Misleading evidence
Forensic evidence
Content analysis
Appeal cases

ABSTRACT

Evidence has the potential to be misleading if its value when expressing beliefs in hypotheses is not fully understood or presented. Although the knowledge base to understand uncertainties is growing, a challenge remains to prioritise research and to continuously assess the magnitude and consequences of misleading evidence in criminal cases. This study used a systematic content analysis to identify misleading evidence, drawing information from case transcripts of rulings argued unsafe by the Court of Appeal of England and Wales. In the 7-year study period, 218 applications were successful on appeal, containing 235 cases of misleading evidence. The majority (76%) of successful appeals were based upon the same materials available in the original trial, rather than the presentation of new relevant information. Witness (39%), forensic (32%), and character evidence (19%) were the most commonly observed evidence types, with the validity of witnesses (26%), probative value of forensic evidence (12%), and relevance of character evidence (10%) being the most prevalent combinations of identified issues. Additionally, the majority (66%) of misleading evidence types relate to their interpretation at activity level. The findings suggest that many of these misleading aspects could have been prevented by providing more transparency in the relationship between evidence and hypotheses. Generally, the results contribute to gaining a more complete picture of the role of misleading evidence in the criminal justice system.

1. Introduction

Uncertainty is an inherent characteristic of many forensic analysis and interpretation processes. Therefore, advancing the role that the growing forensic knowledge base can play throughout the criminal justice system requires an understanding of how uncertainties are currently being dealt with and whether this is improving over time. This coincides with the significant number of concerns discussed in various government-led reports, highlighting the need for research on the scientific validity of methods, judgements, and presentation methods in court [1–3]. As will be discussed in the following sections, empirical research has increasingly focused on these topics, either by testing the validity of methods in experimental settings or by studying processes in mainstream casework, although a lack of studies highlighting the nature and significance of problematic evidence remains [4–7]. More specifically, the UK government has stated:

“It recognises the potential value of the proposed reliability test in reducing the risk of unsafe convictions arising from unreliable expert evidence. However, there is no robust estimate of the size of the problem to be tackled – either in terms of the number of cases where unreliable

expert evidence is adduced, nor in the impact this has in terms of subsequently quashed convictions.” [8]

This study begins to address this gap by systematically assessing the nature, extent, and consequences of ‘unreliable expert evidence’ in legal rulings in England and Wales, by studying the wider issue of misleading evidence within any ruling overturned by the Court of Appeal. More generally, it presents a method which is applicable to other legislations and of which the results can be used to not only develop methods to avoid evidence being misleading in the future, but also to identify possible cases in which it has not surfaced yet.

1.1. ‘Unreliable expert evidence’ and uncertainties

When forensic evidence is used throughout a criminal investigation, it is assumed to have some relation to the criminal act, and therefore has some ability to support the reconstruction of related events. More specifically, analysing an item of evidence aims to determine the value of parameters of this observed evidence (e.g. the refractive index of glass) which can subsequently be used in the interpretation stage to express a belief in hypotheses (e.g. possible sources). However,

* Corresponding author.

E-mail addresses: nadine.smit@ucl.ac.uk (N.M. Smit), ruth.morgan@ucl.ac.uk (R.M. Morgan), d.lagnado@ucl.ac.uk (D.A. Lagnado).

variation may exist between the true and observed parameter value, impacting subsequent interpretations [9,10]. Such variation depends upon the method's accuracy or systematic error (determined by the specificity (true negative rate) and sensitivity (true positive rate), together with a threshold above which the method can be called 'reliable' [11]), as well as on information on its precision or random error (their repeatability and reproducibility) [3,12]. In addition, factors may influence parameter values post-event, such as environmental conditions and collection strategies [13,14]. 'Unreliable expert evidence' then relates to the extent to which the meaning of the uncertainty caused by such factors are and can be considered in the interpretation and presentation stages of the forensic science process.

A growth in empirical research allows for greater understanding of such uncertainties. For example, studies have highlighted factors involved in the dispersion, transfer, and persistence of many different trace evidence types including GSR and trace DNA [15,16]. Additionally, a growing body of research has focussed on sources of potential bias in the analysis processes of domains which rely upon methods with subjective aspects, such as handwriting analysis [17] and the analysis of skeletal remains [18]. Moreover, in addition to the growing analysis of error rates using past cases [19–21] a recent shift can be observed, integrating blind testing programmes within mainstream case examinations [22]. Current interpretation and presentation processes have been the focus of research after identification of misinterpretations of the frequency of features in populations, including those of fingerprints [23,24], hair [25], and bullet-lead [26], and several high profile cases triggered research into the misleading aspects of statistical evidence [27,28]. However, despite the growing knowledge base, uncertainties will remain and so does the need to acknowledge these and understand their effects.

1.2. Misleading evidence and unsafe rulings

Failing to acknowledge some of the discussed uncertainties that are associated with the evaluation of hypotheses (either due to a lack of knowledge or a misinterpretation) may result in an erroneous understanding of the evidential value of evidence, which, if sufficiently significant, results in "reporting support for a hypothesis that is not true" [22] while reporting opposition for a hypothesis that is true (to avoid misleading interpretations where evidence could also support an hypothesis that is true (e.g. a partial mark has common features with multiple prints [29])). Within a legal setting, rulings have been argued unsafe following the identification of issues which in retrospect could have changed the decision of the Trier-of-fact (Criminal Appeal Act 1995). The term 'misleading evidence' will be used here to simply reflect any evidence presented which misled or had the potential to mislead decision-makers in the truth-finding process.

Many studies into unsafe rulings focus on individual or small sets of high-profile cases [30], such as those in England and Wales [31], the Netherlands [32], and Switzerland [33], often presenting specific recommendations following the outcome of each case. Some of the earliest more extensive empirical-based studies have reported issues related to eyewitnesses, informants, bad character, fingerprint forgeries, faked autopsies, and an inadequate defence [34–40]. Although it has been argued that knowledge is still lacking to draw strong inferences about the relationship between forensic evidence and wrongful rulings [8,11], many of the more recent published studies on wrongful convictions include sections on the role of forensic evidence. An overview has been compiled for this study, see Table 1. Public attention rose in the USA by post-conviction DNA testing through the Innocence Project [41], which has been referred to as the beginning of "the age of innocence" [42]. This led to both the acceptance that wrongful convictions happen, as well as a growth in research identifying its causes and developing reforms [43] through the use of more solid data sets [11]. In many of the studies in Table 1, non-DNA evidence was initially used to narrow down the pool of suspects in what was later argued to be

beyond what was scientifically valid [36]. In the decade that followed, studies aimed at highlighting the general severity of problems with forensic evidence [6,44], while others aimed more specifically in categorising these issues [45,46], fuelling the debate on who to blame for these consequences; bad lawyering [47] or bad forensic science [45]. More recently, comparative studies have been performed, recognising 'forensic error' [48] and the amount of evidence types at the trial [49] as predicting factors of unsafe convictions. What the results presented in Table 1 mostly show is that wrongful convictions are not always just an issue of flawed science or bad lawyering, but rather, flawed communication and interpretation [45,47], an issue both sides should take responsibility for.

1.3. Present study: a structural approach in understanding misleading evidence

The idea that evidence has the potential to be misleading throughout an investigation and in court is represented in Fig. 1. This study aims to contribute to understanding the nature of misleading evidence, by presenting and implementing a systematic content analysis, to comprehensively infer the nature of misleading evidence from concerns expressed by appeal judges on the safety of trial rulings (the overlapping area in Fig. 1) in the Court of Appeal of England and Wales. The results will provide insights in both the type of evidence that is misleading (e.g. witness statements or DNA) as well as the nature of this misleading evidence (e.g. the evidence was not reliable or relevant). Of specific interest is whether the issues could have been known and subsequently could have been avoided in the trial. A greater understanding can drive subsequent research, allowing us to draw more general inferences about the complete set of misleading evidence in Fig. 1, to include those who have not been recognised and to avoid similar issues in the future [30,35,59].

For the purpose of the study, we assume that the 'correct' rulings are those based upon the judgement of the appeal court judges, as the ground-truth cannot be known other than by running test-cases through the system. Additionally, allowing for the possibility that there are unsafe rulings which have not been recognised as such [11,30], the results can be used as a (highly liberal) snapshot of the current situation (assuming that rendering a ruling as unsafe is generally justified). Moreover, although it has been suggested that the frequency of (known or revised) errors of justice is relatively low compared to all convictions [30], the disutility of even a single wrongful ruling is extremely high.

2. Materials and methods

Previous research suggests that valuable information can be drawn from structural research on the outcomes of casework [9]. In order to structurally study case documents of wrongful rulings to make valid inferences on the underlying themes, a content analysis approach was used [60,61]. The steps that were undertaken include a systematic case selection, case coding, testing of coding reliability, and analysis of results [62].

2.1. Case selection

The cases used in this study are a convenience sample of all relevant cases, accepting that this only includes misleading evidence identified through the used overturned rulings (see Fig. 1) [60]. Although it has been argued that there is a lack of information on appeal outcomes [63], exacerbated after the discontinuation of Casetrack, appeal decisions were gathered from the Westlaw UK database. They were selected from the *case analysis documents* on the basis of having been heard by the Criminal Court of Appeal of England and Wales (EWCA Crim, further referenced as 'AC'), and having been labelled with *criminal evidence* according to Sweet & Maxwell's Legal Taxonomy [64]. The dataset was further limited to appeals allowed with regards to the conviction or

Table 1
An overview of studies into larger datasets and significant cases where issues with the forensic evidence were observed. It highlights similar datasets (superscripts), number of cases (N) and those containing forensic evidence (F), the percentage of post-conviction DNA tests, and unknowns (u).

Reference	Where	Data	N (F)	PC DNA	Categories of the issues with the forensic evidence
Borchardt & Lutz (1932) [35]	USA (62)/ UK (3)	Significant cases	65 (8)	None	Wrongful conclusions (including handwriting, blood, and ballistics).
Bedau & Radelet (1987) [50]	USA	Death penalty cases: various including NYT index and public inquests ^a	350 (46)	None	Erroneous diagnosis of the cause of death (16) and misleading circumstantial evidence (30)
Rattner (1988) [51]	USA	Books, documents, newspapers	205 (3)	None	Forensic science errors (3)
Wilson (1989) [52]	USA	Newspaper and criminological literature ^b	20 (12)	None	Forensic expert as advocate (5), inconclusive expert evidence (5), circumstantial/suspect evidence (6)
Radelet et al. (1992) [53]	USA	Death penalty cases ^a	416 (u)	None	Corrupt practices, rushing to judgments (unreliability in laboratory work).
Connors et al. (1996) [36]	USA	Various searches	28 (22)	All	Too much weight on non-exclusion and scientific strength of non-DNA evidence (blood, semen, hair).
Sheck et al. (2000) [41]	USA	Innocence project ^c	62 (u)	All	Serology inclusion (32), defective or fraudulent science (21), unreliability, including microscopic hair comparison (18), other forensic inclusions (5), DNA inclusions (1)
Saks & Koehler (2005) [6]	USA	Innocence project ^c	86 (u)	All	Testing errors (63%), misleading expert testimony (27%)
Gross et al. (2005) [44]	USA	Combination of media and website datasets ^d	340 (24)	144 (42%)	Perjury by a forensic scientist
Langdon & Wilson (2005) [54]	AUS/NZ	Law databases, newspaper, books ^b	32 (21)	Partly	Partisan expert testimony (22%), inconclusive expert evidence (31%), circumstantial/suspect evidence (44%)
Roach & Trotter (2005) [31]	UK	IRA Cases	4 (4)	None	Over-estimation of probative value of test result for hypotheses (explosives).
Garrett (2008) [46]	USA	First 200 DNA exonerations ^e	200 (113)	All	- Use of evidence with limited probative value (serology, hair)
Garrett & Neufeld (2009) [55]	USA	DNA exonerations ^e	137 (82)	All	- Improper/misleading testimony by examiners (hair, fibres, bite marks)
					1. "invalid forensic testimony" (82)
					1.1. The misuse of empirical population data
					1.1.1 Non-probative evidence presented as probative (48)
					1.1.2 Exculpatory evidence discounted (23)
					1.1.3 Inaccurate frequency or statistic presented (13)
					1.2. Conclusions on probative value unsupported by empirical data.
					1.2.1. Statistics without empirical support (5)
					1.2.2. Non-numerical statements without empirical support (19)
					1.2.3. Conclusion that evidence originated from defendant (6)
					2. Withholding exculpatory forensic evidence (3)
Collins & Jarvis (2009) [47]	USA	First 200 DNA exonerations ^e	200 (32)	All	Forensic science malpractice
Garrett (2011) [45]	USA	First 250 DNA exonerations ^e	250 (155)	All	- Unreliable forensic evidence (method that does not produce consistent or accurate results). - Invalid conclusions following the analyses. See [55].
Judicial Commission of NSW (2011) [56]	AUS	Conviction appeals in NSW	315 (u)	u	General categories included unreasonable or unsupported jury verdict, issues with the admission/rejection of evidence, and issues with the direction of the judge towards the jury.
Gross & Shaffer (2012) [30]	USA	National registry of exonerations ^d	873 (210)	325 (37%)	False or misleading forensic evidence, ranging from simple mistakes to invalid techniques to fraud
Gould et al. (2013) [48]	USA	Various searches and previous research	260 (88)	Partly	Errors more in testimony or interpretation rather than scientific testing: - Neglecting to provide the jury with key information - Overstating the inculpatory nature by providing inaccurate or non-existent statistics (e.g. hair) - Misstating the certainty of results when the forensic technique does not allow for it. - Poor communication between lab and police and prosecutor - Inadequate training on criminal justice officials - Police unaware of significance, evidence was not revealed - Tunnel vision: more a confirmatory role in the investigation
Garrett (2015) [57]	USA	First 330 DNA exonerations ^e	330 (234)	All	Concealed exculpatory evidence, erroneous analyses, vague testimony of similarity
Dioso-Villa (2015) [58]	AUS	Ex gratia application cases	71 (22)	u	Forensic error or misleading forensic evidence

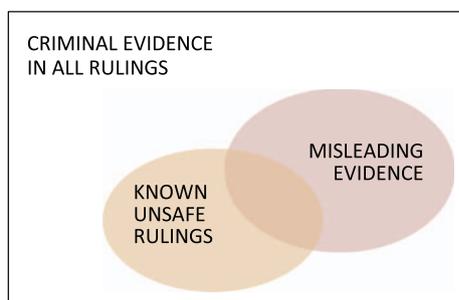


Fig. 1. Diagram showing the present study sample (overlap), a sub-sample of all misleading evidence, obtained through studying overturned unsafe rulings.

acquittal (rather than the sentencing), as this was believed to provide more information on significant misleading evidence, following a belief by appeal judges that the trial ruling was unsafe (outlined in the Criminal Appeal Act 1995). The study term (of the appeal rulings) was a 7-year period January 2010 through December 2016.

2.2. Coding categories and considerations

This study aimed to identify the basis for the successful appeal (i.e. the reason why the trial conviction was unsafe), which is of a descriptive nature and reflects simply the statement of the appeal judge (In Vivo) rather than a normative evaluation of whether that reason was justified. Coding categories were determined using three different methods, combining deductive and inductive techniques. Firstly, codes were developed a priori based upon the results from previous studies presented in Table 1 and the admissibility and criteria of evidence from the Criminal Justice Act 2003 and the Police and Criminal Evidence Act 1984. Examples include issues with the validity of collection and analysis methods, and the determination of the relevance and probative value of evidence. Secondly, the coding framework was adjusted to categorise the issues based upon the possibility of developing methods to further understand and avoid these issues in the future. The most important aspect was differentiating between whether there was new evidence, whether the trial evidence had been re-evaluated by an expert, or whether the decision had been overturned without new evidence. Lastly, the framework was refined through the identification of inductively defined codes from analysing the case documents [61,62]. Added categories include, for example, procedural errors, inconsistent decisions by the jury, and insufficient directions by the judge on the relevance of evidence. The final coding scheme is reflected in the decision scheme in Fig. 2. The following terminology reflects some of the detailed coding categories:

- Issues with the *relevance* of the evidence have to do with the question whether it can be of value to address the hypothesis (e.g. and therefore should have been excluded/included).
- When considering the *probative value* of evidence, it is agreed that the evidence was relevant but it was misleading with respect to how much/little weight was given to the evidence.
- The *validity* of the evidence is used to categorise evidence which was argued by the appeal judge to not be (presented) accurately or precisely.

2.3. Case attributes

Together with the nature and type of the (misleading) evidence, attributes such as the appeal outcome, requests for re-trials, involvement of the Criminal Case Review Commission (CCRC), trial dates, and proposition levels were also considered.

2.4. Inter-coder reliability

To determine the reliability of the coding method, an inter-coder agreement index was developed based upon the coding of the nominal categories of the initial and two additional independent coders (as the initial coder was involved in the development of the coding rules which could impact the reliability [65]), by determining Krippendorff's alpha coefficient [60]. The additional coders were given 20% of the total of cases studied and only relied upon the coding instructions presented in Fig. 2 and the appeal court transcripts. They were required to identify the type of evidence questioned in the trial (and new evidence in the appeal if present), and the nature of why it was misleading (the chosen category following the coding procedure).

3. Results and discussion

3.1. Significance: number of cases

In the 7-year period between 2010 and 2016, 10,859 cases were found in Westlaw UK as having been heard in the Court of Appeal of England and Wales. Of the 996 cases which involved *criminal evidence*, rulings in 218 cases (22%) were argued unsafe because they contained *misleading evidence*, representing the shaded area in Fig. 1. Details of these cases are shown in Table 2. As argued previously by others [11,47,48], this study also suggests that the actual occurrence of misleading evidence is higher. It is possible that many cases are not included in the study sample, either because they have not gone into appeal or because, although errors are identified, they were not argued to “undermine the safety of the conviction” (e.g. see [2015] AC 1619 or [2016] AC 4, or where evidence was argued in the light of new evidence, to be ‘neutral’, see [2010] AC 2936). Logically, however, understanding these issues is of great importance as such errors could potentially contribute to unsafe rulings in other contexts.

Despite the number of cases of misleading evidence found in this study, the challenge with the method developed for this study as compared to most of those discussed in Table 1 (which study already known unsafe rulings), is that only approximately 1 in every 5 in-depth studied cases (22%) heard in the Court of Appeal is rendered unsafe. Because of this, studying all unsafe cases or all cases where forensic evidence is present to determine desired rates in Fig. 1 is not feasible. Such results would require the need for more automated approaches of analysing the large number of case material – something that can only be explored once a more detailed picture exists of misleading evidence in case transcripts as produced for this initial study.

3.2. Further significance: time between rulings

The significance of the 223 overturned counts becomes more pertinent when considering the time between the original conviction and the decision by the appeal judges. This time difference in cases where this could be obtained ($n = 208$) averaged 2.9 years but with a wide range ($sd = 5.7$ years). Eighteen rulings took over 10 years to be overturned, with a maximum time difference of 36.3 years (in [2014] AC 2047, which was quashed due to fresh evidence indicating police malpractice). Moreover, the method of data collection allowed for more detailed analyses. For example, comparing the median of the time difference (interquartile range) of 2.7 (1.3 – 9.7) years for rulings where new evidence had to be obtained with the 0.8 (0.5 – 1.3) years for rulings which were overturned without having to acquire new evidence, shows a significant difference between the two distributions (Mann–Whitney $U = 5684$, $n_{new} = 47$, $n_{notnew} = 179$, $p < 0.001$, two-tailed). This is consistent with the length of time it would take to acquire and consider new evidence in a case.

It followed from the case documents that of the 201 overturned convictions, no re-trial was ordered in 80 cases (in addition to 24 re-trials and 97 unknowns). Additionally, re-trials were requested

Is the ruling regarded unsafe because there is new material available in the appeal not known during the trial? (also incl. re-evaluation of the trial evidence)

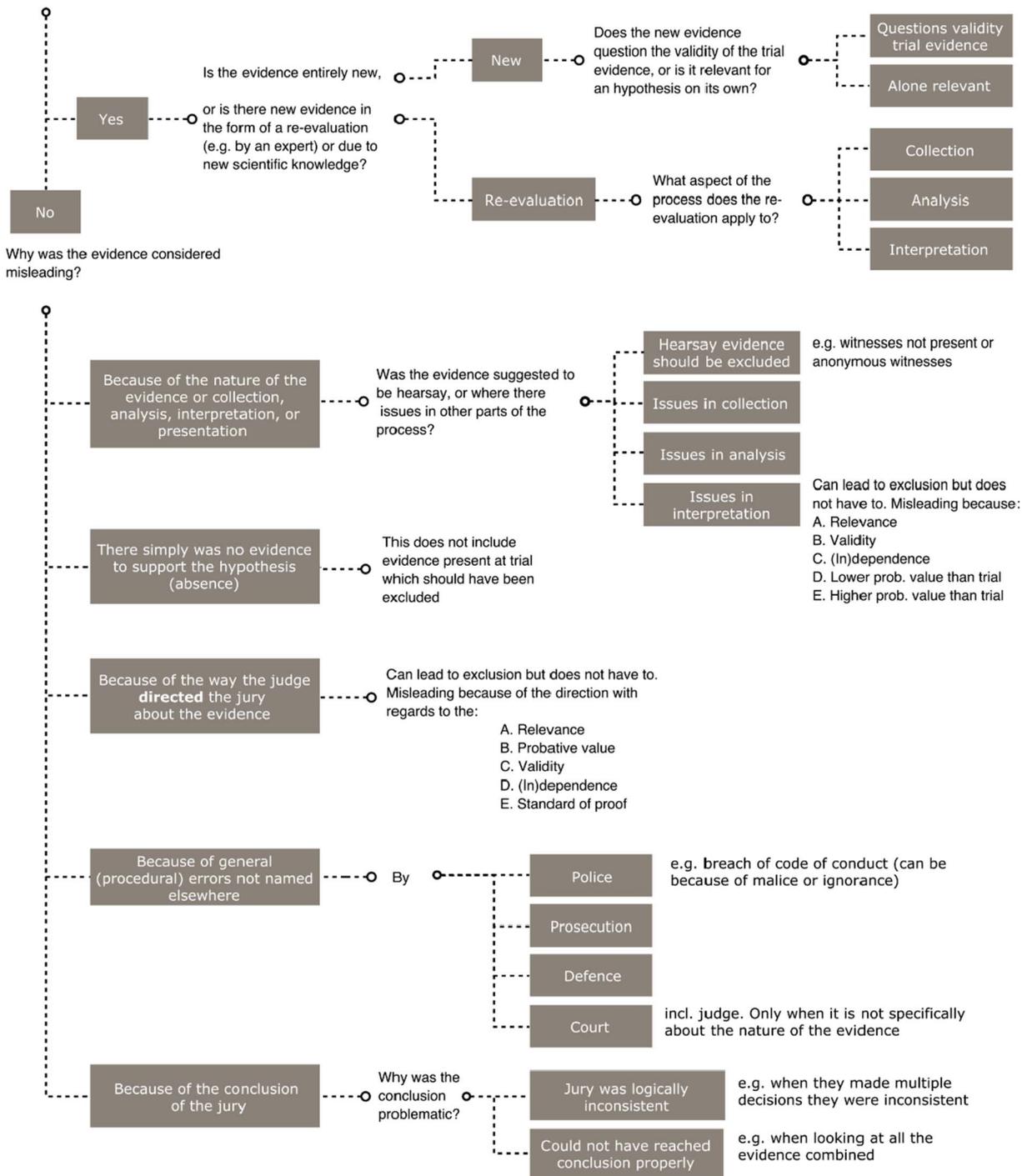


Fig. 2. The coding scheme used in this study to classify the reasons for rendering the trial ruling unsafe by the Court of Appeal judge(s).

following all 10 acquittals (including resumed proceedings), a possibility after the double jeopardy ban in the UK in 2003 (Criminal Justice Act 2003), see for example [2012] AC 414. In 10 cases, no re-trial was ordered as the conviction was substituted (e.g. murder to manslaughter). More generally, the time between the rulings does not always correspond with the time spent incarcerated, as, for example, the suspect may have been incarcerated before the trial date or have already completed their sentence before the appeal. The cases involved significant crimes, including those related to sexual assault ($n = 50$), assault (34), murder (28), robbery (18), fraud (13), and drug offences

(13). Nevertheless, whether a suspect was ultimately guilty or not does not change the fact that misleading evidence was still present in these trials in the first place.

3.3. Coding method reliability

The coding method reliability has been studied by determining the inter-coder agreement on 20% of the total number of cases coded for this study, resulting in a Krippendorff's alpha coefficient of 0.772 (47 subjects, 3 coders). This has been suggested as 'substantial' agreement

Table 2
Details of the 218 successful appeal cases between 2010 – 2016 as extracted from Westlaw UK. A total of 996 relevant cases were studied.

	N
Successful appeal cases	218
Rulings (suspects)	222
Conviction quashed	201
Conviction substituted	10
Acquittal quashed	10
Admission of evidence allowed	1
Counts	223
Items of evidence	230
Reasons for misleading evidence	235

Table 3
Overview of the results of each coding category from Fig. 2 representing the reason for the unsafe ruling.

Coding category representing the reason for unsafe ruling	n (%)
a. Misleading nature of the evidence as presented in court	85 (36%)
Relevance	31
Probative value	18
Validity	17
Hearsay	16
Independence	2
Collection	1
b. Issues with the direction of the judge on	60 (26%)
Probative value	18
Standard of proof	17
Relevance	14
Validity	10
Independence	1
c. There is new relevant information	56 (24%)
New questions validity trial	26
New is relevant stand-alone	19
Expert re-evaluation of probative value	7
New questions probative value trial	3
Expert re-evaluation of collection method	1
d. General errors by prosecution (6), police (2), or court (10)	18 (8%)
e. Absence of evidence in the trial	9 (4%)
f. Issues with the decision of the jury	7 (3%)
	235

[66], although others argue that above 0.8 would be considered reliable [60]. The coefficient presented here is, however, a conservative agreement as all categories are considered equally different, while issues with, for example, the probative value or the relevance are somewhat similar and not always easy to distinguish from the case transcript. This does indicate that perhaps a more detailed explanation of the coding categories and/or the reasoning of the judge could be valuable.

3.4. Overview of the results in each coding category

Overtures of 213 of the 223 counts were based upon a single key reason related to a single evidence type, highlighting the importance of avoiding any type of misleading evidence in trial cases. The major and minor categories in which the reason for the unsafe ruling was classified using the coding method (see Fig. 2) can be found in Table 3. It follows that in a minority of instances (24%) there was entirely new evidence or newly gained knowledge (such as a re-evaluation by a forensic expert) which led to the unsafe ruling. In the majority (76%) of cases, however, the ruling was overturned given the same set of information available in the trial. These results do strongly suggest that there might be many more unsafe rulings beyond the post-conviction DNA testing cases summarised in Table 1. Additionally, from the 22 cases which were successful on appeal following a reference by the CCRC (according to the transcripts), the majority (15 cases) were rendered unsafe due to the introduction of new evidence. This, again, suggests that perhaps not

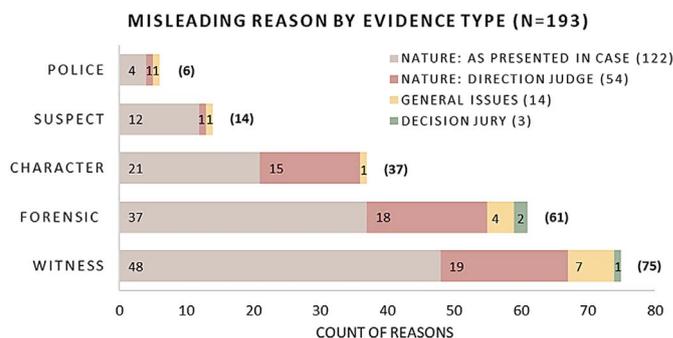


Fig. 3. An overview of the evidence type groups and the reasons given why the evidence was misleading. 42 reasons did not specifically apply to an evidence type and are not included in this graph.

enough focus is directed on cases where there is not necessarily new evidence.

The reasons for why the evidence is misleading are studied from the point of the trial evidence. On the 19 occasions where there was new evidence which was relevant on its own (mostly witnesses (n = 4), documents (3), medical evidence (3), and DNA (2)) and the 9 cases where there was an absence of evidence, no trial evidence was questioned. Additionally, there was no specific evidence associated with the unsafe ruling in some instances where there were issues with the direction given by the judge (n = 6), general errors (4), and issues with the decision of the jury (4). The remaining 193 reasons related to trial evidence are studied further. Here, the remaining cases where there was new relevant information questioning the trial evidence (c in Table 3) are considered together with the nature of the evidence as presented in court (a) as they both relate to the nature of the trial evidence, and only differ in the way the issue came to light (n = 122). Moreover, the issues with the direction of the judge (b) indicate a misunderstanding of this similar nature of the evidence (n = 54), an issue also identified by others [56]. The presence of these revised categories is studied with respect to the evidence type groups shown in Fig. 3.

It follows from Fig. 3 that witness (including victim) statements are the most problematic evidence type (39%), which has also been suggested as a major cause for concern by other scholars [30,50,57]. The second most-common evidence type is the grouped category of forensic evidence (32%), which will be a major focus of the subsequent sections. Although discussions have been ongoing as to whether (good and bad) character evidence (19%) has a higher risk of convicting the innocent and/or whether the underlying assumptions are fair towards a suspect [67], especially as after becoming more easily admissible after the Criminal Justice Act 2003, research did not find evidence of unfair prejudice [68]. The final two evidence types are police- and suspect statements. While false confessions have been identified as a significant factor in wrongful convictions in the USA [30] and can generally have a major impact in an investigation [69], this study has only found four such cases, and are therefore either less common or were not identified in the present study. The four general reasons in Fig. 3 are grouped in two sets and discussed in Sections 3.5 and 3.6.

3.5. Issues with the nature of the evidence

The two major categories related to the nature of the evidence are responsible for the majority of the overturned reasons found in this study (n = 176 or 75%), see Fig. 3. In other words, the common theme is that of possible erroneous beliefs in hypotheses in the light of such misleading evidence. In order to improve this process, it is necessary to understand what type of hypotheses are addressed with this evidence. This allows for the possibility to develop further knowledge bases with respect to each evidence type, such as within- and between source characteristics and evidence dynamics. The findings with respect to the

Table 4
Overview of the number of cases of misleading evidence and the type of hypothesis that they addressed.

	Hypothesis related to				
	Activity	Source	Validity	Intent	
Witnesses evidence	48	29	6	2	85
Police	2	3			5
Suspect	9	1	2	1	13
Other witness	37	25	4	1	67
Forensic evidence	39	10	4	1	55
Animal blood	1				1
Digital	3	1			4
DNA	4	4			8
Documents	11		1		12
Drugs	5		1	1	7
Fingerprints	4	2			6
Footwear marks		1			1
Glass		1			1
GSR	1				1
Handwriting		2			2
Medical	7		1		8
Weapon	3		1		4
Character evidence	29	2	5		36
	116	42	15	3	176

hypotheses in question in each case are shown in Table 4.

It follows from the findings presented in Table 4 that the majority (116 or 66%) of items of evidence were used to address hypotheses related to an activity level, such as witnesses describing violent acts of suspects, victims describing sexual offences by suspects, or cases where evidence of bad character was used as evidence that a suspect was more likely to commit certain acts. The results highlight the importance of understanding trace evidence dynamics and the value and limitations of observing evidence such as DNA, fingerprints, and GSR when addressing these activity level hypotheses. Moreover, cases where there was a dispute about the lower, source level, hypotheses, were also commonly encountered whereby the evidence of witnesses and forensic evidence were misleading. In the remaining cases, the evidence was used to assess the validity of other evidence (this is not the same as issues with the validity in Table 3) and used to determine whether or not intent was involved.

The final step in moving forward and addressing these issues by developing and better employing the forensic knowledge base is the

Table 5
Overview of the number of cases of misleading evidence and the detailed relationship between the different levels of hypotheses (see Table 4) and the nature of the misleading evidence, grouped by general evidence category.

	Misleading nature				
	Relevance	Probative value	Validity	Standard of proof	
Witness evidence	8	14	60	3	85
Activity	4	7	35	2	
Source		6	22	1	
Validity	3	1	2		
Intent	1		1		
Forensic evidence	13	29	7	6	55
Activity	10	19	5	5	
Source	1	8	1	1	
Validity	2	1	1		
Intent		1			
Character evidence	24	6	4	2	36
Activity	20	4	3	2	
Source	1	1			
Validity	3	1	1		
	45	49	71	11	176

need to understand the misleading aspect of the relationship between the evidence and the hypothesis. For example, was the evidence simply not relevant for the hypothesis? Was it given too much evidential weight? Or was the evidence itself not valid and therefore problematic for assessing the truth in the hypothesis? These results (using the sub-categories in Table 3) are shown in Table 5, grouped per general evidence category and hypothesis level (from Table 4). It follows from Table 3 that 16 cases were categorised as hearsay, following, for example, the argument that because a witness could not be cross-examined, the possibility of assessing their reliability was severely diminished ([2012] AC 1509). These cases were therefore categorised as questions on the (general) validity of the evidence. The three cases categorised under *issues with the independence of evidence* were rendered unsafe because of underlying issues of validity (whereby an identification was influenced by possible biasing information) and probative value (where two dependent items of evidence were presented as providing independent support for an hypothesis).

3.5.1. Relevance of evidence

First of all, the evidence must be relevant for hypotheses (Police and Criminal Evidence Act 1984). With regards to the character evidence, this assessment was the major issue, mainly whereby bad character was ‘unfairly’ used as evidence that a suspect could have committed a criminal act. Evidence deemed irrelevant for the hypothesis in question was also observed in relation to several forensic evidence types, including cases where the presence of drugs was arguably wrongly used as evidence to support hypothesis in relation to handling or dealing (other types of) drugs (e.g. [2013] AC 901 and [2012] AC 2879).

In order to avoid such misinterpretations, it is crucial to define sub-hypotheses which are perhaps more clearly associated with the recovered evidence to provide transparency in the reasoning process. For example, when the presence of marijuana is used as evidence for dealing heroin, a sub-hypothesis can be defined on the dealing of marijuana, and how this is linked with the presence of marijuana on the one hand, and the dealing of heroin on the other. These relationships can then be assessed separately but interpreted as a whole.

3.5.2. Probative value of evidence

After evidence is considered relevant for an hypothesis, the belief expressed in an hypothesis (with respect to another) is sometimes presented disproportionately to the actual probative value of an item of evidence. This issue was most often related to the forensic evidence (59%), and was also the major issue within the forensic evidence (53%). Examples on the activity level include medical evidence such as trauma, argued to be conclusive of penetration ([2012] AC 1433) or non-accidental trauma ([2010] AC 2847), which were on appeal revised considering new scientific findings. Within trace evidence, the appeal highlighted questions such as those regarding the time when DNA was deposited ([2015] AC1732), to what extent the location of DNA ([2011] AC 460) and fingerprints ([2010] AC 2421) were suggestive of the criminal act, the possibilities of multiple transfer of blood ([2010] AC 2499), and the exact meaning of small quantities of GSR on clothing ([2014] AC 2507). Similarly, the appeal court rendered convictions unsafe where disproportionate weight was given to evidence when addressing questions of identifications or individualisations on a source level. Appeal rulings suggested the need to consider possible sources of glass unconnected to the crime ([2015] AC 1950), the misleading aspect of the trial judge failing to direct the jury about the meaning of a partial DNA profile ([2010] AC 1334), the importance of recognising disputes between fingerprint experts ([2011] AC 1296), and severe limitations in the notorious case of R. v T when determining the source of a shoe print ([2010] AC 2439).

In addition to identifying sub-hypotheses clarifying the relationships between evidence and hypotheses, determining the change in belief in an hypothesis following the evidence requires an in-depth analysis of the factors that can influence the state of the evidence given

the specific hypothesis. As argued previously, it requires an extensive knowledge base on the factors influencing the dynamics of evidence and the abundance of observed characteristics. Such datasets can be generally developed, but it should be made clear under what assumptions these are applied in a case context and how this impacts the interpretation.

3.5.3. Validity of evidence

When considering the witness statements, issues were most often associated with regards to their validity (e.g. the credibility of a witness). Although the validity of evidence is not directly linked to the relationship between evidence and an hypothesis (as compared to the relevance and probative value), a decrease in the validity of the evidence (whether it is the accuracy or precision) can, again, cause a misinterpretation of the probative value or relevance of the evidence (although these latter assessments could be justified themselves).

In relation to the witness statements, and to a lesser extent the character and forensic evidence with a questionable validity (including fingerprints, handwriting evidence, and digital evidence), it is pertinent that the factors influencing the validity are understood. These include, for example, the conditions under which the witness observations have been made (e.g. distance, weather) and the degree of subjectivism of forensic analysis methods and factors which can influence this process (as discussed in the introduction). Judges and jurors should be made aware of these effects and their implications for the belief in hypotheses, as they generally tend to consider forensic evidence as quite reliable [70].

3.5.4. Standard of proof

One of the roles of the judge is to direct the jury on the standard of proof [71]. Part of this belief is an understanding of the factors discussed previously in this section. However, some of the issues do not fit in these categories and are related to the standard of proof more generally. Examples not related to a specific evidence type (see section 3.4) include a lack of direction by the judge on what hypotheses they had to be sure of before they could convict (e.g. [2016] AC 1632). Moreover, the 11 cases where there was specific evidence (Table 5), the direction on the standard of proof related both to source and activity level hypotheses. It was, for example, argued that the judge failed to direct the jury that they could reach their own conclusions on the meaning of low-template DNA evidence ([2010] AC 549), and that the judge did not direct the jury on the law related to circumstantial drug-related evidence ([2015] AC 1733). Overcoming these issues from the point of the interpretation of the forensic evidence will mostly relate to a transparent presentation of the hypotheses the jury needs to assess, the uncertainties of the presented analyses or interpretations, and, together with that, an indication that the jury a) should reach their own conclusions based on these uncertainties and b) the extent to which the jury should be 'sure' of its conclusion.

3.6. General issues and the decisions by the jury

In addition to the nature of the evidence which is (or has the potential to be) misleading in the truth-finding process, it follows from Fig. 3 that in 14 cases there were 'general issues' with the evidence, including procedural errors by the court, failures to disclose statements by the prosecution to the defence, and breaches of code by police officers. These only form a small proportion of all cases and would require a more in-depth analysis of research targeted at these issues. Additionally, three cases were overturned where the appeal judges argued that the decisions of the trial jury were either logically inconsistent or where there was no possibility for them to reach a conclusion properly (e.g. [2010] AC 130). Although it is difficult to evaluate the fairness and reasonability of the decisions of a jury, research has shown a variation

in jurors' interpretation of the evidence following, for example, factors that might influence the validity of eyewitnesses [72], the so-called CSI-effect [73], conditions affecting the interpretation of probabilistic evidence [74,75], and a change in the method of presentation (e.g. qualitative versus quantitative) of identification evidence [76]. Therefore, some of the requirements needed to improve the interpretation of forensic evidence that follow from the results in this paper should coincide with results from juror decision-making studies.

4. Conclusion

Content analysis of appeal court transcripts can provide vast information on the presence and nature of misleading evidence. Although the sampling method causes a danger of over-analysing the data presented in this study, it does suggest that the relevance, probative value, and validity of evidence are often misunderstood and miscommunicated within a criminal trial when expressing beliefs in (competing) hypotheses. The consequences of these are severe and have caused many defendants to be wrongfully incarcerated. The results show that these issues can be addressed by clarifying sub-hypotheses on source and activity levels and by developing the knowledge base on, for example, the rarity of observed characteristics in a population and the factors affecting evidence dynamics. The fact that most of these rulings are overturned without including new evidence in the appeal suggests that they could have been avoided in the trial if these interpretations were improved. Accepting that uncertainty will remain present, it is important that the uncertainties are not caused by erroneous arguments and judgements but by inherent properties of the knowledge base, which, in turn, can be presented transparently.

The findings from this research support previous findings that the identified cases are only the tip of the iceberg and can no longer be attributed to simple individual "bad apples" in the system [77]. This study provides a foundation of detailed issues related to the interpretation of forensic evidence and highlights the possibility of identifying these issues from case transcripts, fuelling the important debate on improving access to case documents post-conviction [78]. On the one hand, the findings from this research can be validated by studying and improving the understanding of uncertainties involved in assessing the relevance, probative value, and validity of evidence on a more comprehensive level for individual evidence types. This also includes the use of methods to guide decision-making under uncertainty and with new information, such as Bayesian networks [79–81]. On the other hand, there is great potential in applying similar methods to larger scale studies within England and Wales as well as in other jurisdictions, using, for example, more automated text-mining techniques. These can also be used to determine potential predicting factors of misleading evidence (see for general examples [48]), using specific factors related to the type of evidence and the type of hypotheses they address. The findings in this paper suggest that, as others have argued, "the scientific community can take the lead in reform efforts" [55]. It is therefore hoped that this research becomes more than just an "academic exercise" as these datasets grow, and that these results allow for the much-needed continuous evaluation of the use of forensic evidence in court and their contribution to both justified and unjustified rulings.

Acknowledgements

This research was funded by the Engineering and Physical Sciences Research Council of the UK through the Security Science Doctoral Research Training Centre (UCL SECRt) based at University College London (EP/G037264/1). The authors would like to thank Sherry Nakhaeizadeh, Helen Earwaker, and Bluebelle Carroll for their help in assessing the reliability of the coding method.

References

- [1] The Law Commission, *Expert Evidence in Criminal Proceedings in England and Wales*, London, (2011).
- [2] National Research Council, *Strengthening Forensic Science in the United States: a Path Forward*, Washington D.C. (2009).
- [3] PCAST, *Forensic Science in Criminal Courts: Ensuring Scientific Validity of Feature-comparison Methods*, (2016).
- [4] S.R. Gross, B. O'Brien, Frequency and predictors of false conviction: why we know so little, and new data on capital cases, *J. Empir. Leg. Stud.* 5 (2008) 927–962, <http://dx.doi.org/10.1111/j.1740-1461.2008.00146.x>.
- [5] S.R. Gross, Convicting the innocent, *Annu. Rev. Law Soc. Sci.* 4 (2008) 173–192, <http://dx.doi.org/10.1146/annurev.lawsocsci.4.110707.172300>.
- [6] M.J. Saks, J.J. Koehler, The coming paradigm shift in forensic identification science, *Science* 309 (2005) 892–895, <http://dx.doi.org/10.1126/science.1111565>.
- [7] J.P. Bono, Commentary on the need for a research culture in the forensic sciences, *UCLA L. Rev.* 58 (2011).
- [8] UK Ministry of Justice, *Government Response to Law Commission Report on Expert Evidence*, (2013).
- [9] J.L. Mnookin, S.A. Cole, I.E. Dror, B.A.J. Fisher, The Need for a Research Culture in the Forensic Sciences, *UCLA L. Rev.* 58 (2011) 725.
- [10] A.M. Christensen, C.M. Crowder, S.D. Ousley, M.M. Houck, Error and its meaning in forensic science, *J. Forensic Sci.* 59 (2014) 123–126, <http://dx.doi.org/10.1111/1556-4029.12275>.
- [11] S.A. Cole, Forensic science and wrongful convictions: from exposé to contributor to corrector, *New Engl. Law Rev.* 46 (2012).
- [12] D.L. Streiner, G.R. Norman, "Precision" and "accuracy": two terms that are neither, *J. Clin. Epidemiol.* 59 (2006) 327–330, <http://dx.doi.org/10.1016/j.jclinepi.2005.09.005>.
- [13] K.R. Scott, R.M. Morgan, V.J. Jones, N.G. Cameron, The transferability of diatoms to clothing and the methods appropriate for their collection and analysis in forensic geoscience, *Forensic Sci. Int.* 241 (2014) 127–137, <http://dx.doi.org/10.1016/j.forsciint.2014.05.011>.
- [14] R.M. Morgan, M. Little, A. Gibson, L. Hicks, S. Dunkerley, P.A. Bull, The preservation of quartz grain surface textures following vehicle fire and their use in forensic enquiry, *Sci. Justice* 48 (2008) 133–140, <http://dx.doi.org/10.1016/j.scijus.2008.04.001>.
- [15] M. Maitre, K.P. Kirkbride, M. Horder, C. Roux, A. Beavis, Current perspectives in the interpretation of gunshot residues in forensic science: a review, *Forensic Sci. Int.* 270 (2017) 1–11, <http://dx.doi.org/10.1016/j.forsciint.2016.09.003>.
- [16] G. Meakin, A. Jamieson, DNA transfer: review and implications for casework, *Forensic Sci. Int. Genet.* 7 (2013) 434–443, <http://dx.doi.org/10.1016/j.fsigen.2013.03.013>.
- [17] B. Found, J. Ganas, The management of domain irrelevant context information in forensic handwriting examination casework, *Sci. Justice* 53 (2013) 154–158.
- [18] S. Nakhaeizadeh, I.E. Dror, R.M. Morgan, Cognitive bias in forensic anthropology: visual assessment of skeletal remains is susceptible to confirmation bias, *Sci. Justice* 54 (2014) 208–214, <http://dx.doi.org/10.1016/j.scijus.2013.11.003>.
- [19] S.A. Cole, More than zero: accounting for error in latent fingerprint identification, *J. Crim. Law Criminol.* 95 (2005) 985–1423.
- [20] M.B. Thompson, J.M. Tange, D.J. McCarthy, Human matching performance of genuine crime scene latent fingerprints, *Law Hum. Behav.* 38 (2014) 84–93, <http://dx.doi.org/10.1037/lhb0000051>.
- [21] A. Kloosterman, M.J. Sjerps, A. Quak, Error rates in forensic DNA analysis: definition, numbers, impact and communication, *Forensic Sci. Int. Genet.* 12 (2014) 77–85, <http://dx.doi.org/10.1016/j.fsigen.2014.04.014>.
- [22] W. Kerkhoff, R.D. Stoel, C.E.H. Berger, E.J.A.T. Mattijssen, R. Hermsen, N. Smits, H.J.J. Hardy, Design and results of an exploratory double blind testing program in firearms examination, *Sci. Justice* 55 (2015) 514–519, <http://dx.doi.org/10.1016/j.scijus.2015.06.007>.
- [23] Office of the Inspector General, *A Review of the FBI's Handling of the Brandon Mayfield Case*, (2006).
- [24] Scottish Criminal Records Office, *McKie Report*, (2006).
- [25] FBI, *FBI Testimony on Microscopic Hair Analysis Contained Errors in at Least 90 Percent of Cases in Ongoing Review*, <https://www.fbi.gov/news/pressrel/press-releases/fbi-testimony-on-microscopic-hair-analysis-contained-errors-in-at-least-90-percent-of-cases-in-ongoing-review>, (2015) (accessed February 13, 2016).
- [26] National Research Council, *Forensic Analysis, Weighing Bullet Lead Evidence*, The National Academies Press, Washington, DC, 2004.
- [27] R. Meester, M. Collins, R. Gill, M. van Lambalgen, On the (ab)use of statistics in the legal case against the nurse Lucia de B, *Law Probab. Risk* 5 (2007) 233–250, <http://dx.doi.org/10.1093/lpr/mgm003>.
- [28] M. Buchanan, Statistics: conviction by numbers, *Nature* 445 (2007) 254–255, <http://dx.doi.org/10.1038/445254a>.
- [29] W.C. Thompson, The potential for error in forensic DNA testing (and how that complicates the use of DNA databases for criminal identification), *Forensic DNA Databases Race Issues, Abus. Actions*, 2008, pp. 1–49.
- [30] S.R. Gross, M. Shaffer, Exonerations in the United States, 1989–2012, *U Michigan Public Law Work. Pap.* 277, 2012, p. 103.
- [31] K. Roach, G. Trotter, Miscarriages of justice in the war against terrorism, *Penn State Law Rev.* 109 (2005).
- [32] C. Brants, Wrongful convictions and inquisitorial process: the case of the Netherlands, *Univ. Cincinnati Law Rev.* 80 (2011) 1069–1114.
- [33] M. Killias, Wrongful convictions in Switzerland: the experience of a continental law country, *Wrongful Convict. Int. Perspect. Miscarriages Justice*, Temple University Press, 2008, pp. 139–156.
- [34] J. Frank, B. Frank, *Not Guilty*, Doubleday, Garden City, New York, 1957.
- [35] E.M. Borchardt, E.R. Lutz, *Convicting the Innocent: Sixty-five Actual Errors of Criminal Justice*, Garden City Publishing Company, Garden City, New York, 1932.
- [36] E. Connors, T. Lundregan, N. Miller, T. McEwen, *Convicted by Juries, Exonerated by Science: Case Studies in the Use of DNA Evidence to Establish Innocence After Trial*, (1996).
- [37] E.D. Radin, *The Innocents*, Morrow, New York, 1964.
- [38] R. Brandon, C. Davies, *Wrongful Imprisonment: Mistaken Convictions and Their Consequences*, (1973).
- [39] B. Geller, J. Almog, P. Margot, E.A. Springer, A chronological review of fingerprint forgery, *J. Forensic Sci.* 44 (1999) 963–968.
- [40] R.L. Fricker, Pathologist's plea adds to turmoil: discovery of possibly hundreds of faked autopsies helps defense challenges, *ABA J.* 79 (1993) 24.
- [41] B. Sheek, P. Neufeld, J. Dwyer, *Actual Innocence: Five Days to Execution, and Other Dispatches From the Wrongly Convicted*, Doubleday, 2000.
- [42] M. Zalman, An integrated justice model of wrongful convictions, *Alb. L. Rev.* 2011, pp. 1465–1525.
- [43] F. Leverick, J. Chalmers, *Causes of wrongful conviction, Post-corroboration Safeguards Rev. Rep. Acad. Expert Gr.*, The Scottish Government, Edinburgh, 2014, pp. 30–43.
- [44] S.R. Gross, K. Jacoby, D.J. Matheson, N. Montgomery, J. Daniel, N. Montgomery, Exonerations in the United States 1989 through 2003, *J. Crim. Law Criminol.* 95 (2005) 523–560.
- [45] B.L. Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong*, (2011).
- [46] B.L. Garrett, *Judging innocence*, *Columbia Law Rev.* 108 (2008) 55–142.
- [47] J.M. Collins, J. Jarvis, The wrongful conviction of forensic science, *Forensic Sci. Policy Manag. An Int. J.* 1 (2009) 17–31, <http://dx.doi.org/10.1080/19409040802624067>.
- [48] J.B. Gould, J. Carrano, R.A. Leo, J.K. Young, Predicting erroneous convictions: a social science approach to miscarriages of justice, *Univ. San Fr. Law Res. Pap.* 2013–20, 2013.
- [49] T.R. Harmon, Too late for luck: a comparison of post-Furman exonerations and executions of the innocent, *Crime Delinq.* 51 (2005) 498–520, <http://dx.doi.org/10.1177/0011128705275977>.
- [50] H. Bedau, M. Radelet, Miscarriages of justice in potentially capital cases, *Stanford Law Rev.* 40 (1987) 21–179, <http://dx.doi.org/10.2307/1228828>.
- [51] A. Rattner, Convicted but innocent - Wrongful conviction and the criminal justice system, *Law Hum. Behav.* 12 (1988) 283–293, <http://dx.doi.org/10.1007/BF01044385>.
- [52] P.R. Wilson, When justice fails: a preliminary examination of serious criminal cases in Australia, *Aust. J. Soc. Issues* 24 (1989).
- [53] M. Radelet, H. Bedau, C. Putnam, *In Spite of Innocence: Erroneous Convictions in Capital Cases*, Northeastern University Press, 1992.
- [54] J. Langdon, P. Wilson, When justice fails: a follow-up examination of serious criminal cases since 1985, *Curr. Issues Crim. Just.* 17 (2005) 179–202.
- [55] B.L. Garrett, P.J. Neufeld, Invalid forensic science testimony and wrongful convictions, *Va. Law Rev.* 95 (2009) 1–97.
- [56] *Judicial Commission of N.S.W., Conviction Appeals in New South Wales*, (2011).
- [57] B.L. Garrett, *Convicting the Innocent Redux*, in: D.S. Medwed (Ed.), *Wrongful Convictions DNA Revolut*, Cambridge University Press, Cambridge, 2015, pp. 40–56, <http://dx.doi.org/10.1017/9781316417119.004>.
- [58] R. Dioso-Villa, A repository of wrongful convictions in Australia: first steps towards estimating prevalence and causal contributing factors, *Flinders Law J.* 17 (2015) 163–202.
- [59] D.K. Rossmo, Case rethinking: a protocol for reviewing criminal investigations, *Police Pract. Res.* 4263 (2014) 1–17, <http://dx.doi.org/10.1080/15614263.2014.978320>.
- [60] K. Krippendorff, *Content Analysis: An Introduction to its Methodology*, SAGE, 2007.
- [61] J. Drisko, T. Maschi, *Content Analysis*, Oxford University Press, 2015, <http://dx.doi.org/10.1093/acprof:oso/9780190215491.001.0001>.
- [62] M.A. Hall, R.F. Wright, Systematic content analysis of judicial opinions, *Calif. Law Rev.* 96 (2008) 63–122, <http://dx.doi.org/10.2307/20439171>.
- [63] J. Jacobson, *By Mistakes We Learn? — A Review of Criminal Appeals Against Sentence*, (2013).
- [64] M. Scott, N. Smith, Legal taxonomy from Sweet & Maxwell, *Leg. Inf. Manag.* 10 (2010) 217–222, <http://dx.doi.org/10.1017/S1472669610000216>.
- [65] M. Lorr, D. McNair, Methods relating to evaluation of therapeutic outcome, *Methods Res. Psychother.* (1966).
- [66] J. Landis, G. Koch, The measurement of observer agreement for categorical data, *Biometrics* (1977).
- [67] J.R. Spencer, *Evidence of Bad Character*, 3rd ed., Hart Publishing, 2016, <http://dx.doi.org/10.5040/9781509900077>.
- [68] M. Redmayne, The relevance of bad character, *Camb. Law J.* 61 (2002) 684–714, <http://dx.doi.org/10.1017/S0008197302001770>.
- [69] S.M. Kassir, Why confessions trump innocence, *Am. Psychol.* 67 (2012) 431–445, <http://dx.doi.org/10.1037/a0028212>.
- [70] B.L. Garrett, G. Mitchell, Forensics and fallibility: comparing the views of lawyers and judges, *West VA. Law Rev.* 119 (2016) 2016–2063.
- [71] *Judicial College, Crown Court Compendium*, (2017).
- [72] R.S. Schmechel, T.P. O'Toole, C. Easterly, E.F. Loftus, Beyond the Ken? Testing Jurors' understanding of eyewitness reliability evidence, *Jurimetrics* 46 (2006) 177–214, <http://dx.doi.org/10.2307/29762929>.
- [73] N.J. Schweitzer, M.J. Saks, The CSI effect: popular fiction about forensic science affects the public's expectations about real forensic science, *Jurimetrics* 47 (2007)

- 357–364, <http://dx.doi.org/10.2307/29762978>.
- [74] W. Thompson, Are juries competent to evaluate statistical evidence? *Law Contemp. Probl.* (1989).
- [75] D.H. Kaye, J.J. Koehler, Can Jurors understand probabilistic evidence? *J. R. Stat. Soc. Ser. A* 154 (1991) 75, <http://dx.doi.org/10.2307/2982696>.
- [76] D. McQuiston-Surrett, M.J. Saks, The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Factfinders Hear, (2009), <http://dx.doi.org/10.1007/s10979-008-9169-1>.
- [77] W.C. Thompson, Beyond bad apples: analyzing the role of forensic science in wrongful convictions, *Southwest. Univ. Law Rev.* 1 (2010) 971–994.
- [78] M. Henneberg, Worlds Apart: cold case reviews and investigations into alleged wrongful convictions in England and Wales, *J. Cold Case Rev.* (2017) 3.
- [79] N.M. Smit, D.A. Lagnado, R.M. Morgan, N.E. Fenton, Using Bayesian networks to guide the assessment of new evidence in an appeal case, *Crime Sci.* 5 (2016) 9, <http://dx.doi.org/10.1186/s40163-016-0057-6>.
- [80] P. Garbolino, F. Taroni, Evaluation of scientific evidence using Bayesian networks, *Forensic Sci. Int.* 125 (2002) 149–155, [http://dx.doi.org/10.1016/S0379-0738\(01\)00642-9](http://dx.doi.org/10.1016/S0379-0738(01)00642-9).
- [81] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, 2nd ed., Morgan Kaufmann Publishers Inc., San Francisco, CA, 1988, <http://dx.doi.org/10.2307/2026705>.