Science & Justice xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

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

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, 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

* Corresponding author.
E-mail addresses: nadine.smit.12@ucl.ac.uk (N.M. Smit), ruth.morgan@ucl.ac.uk (R.M. Morgan), d.lagnado@ucl.ac.uk (D.A. Lagnado).

http://dx.doi.org/10.1016/j.scijus.2017.09.005
Received 30 August 2017; Received in revised form 21 September 2017; Accepted 23 September 2017

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 lead 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 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