Science and Justice 49 (2009) 298–308

Contents lists available at ScienceDirect

Science and Justice

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

Forensic voice comparison and the paradigm shift ☆

Geoffrey Stewart Morrison [a,b]

[a] School of Language Studies, Australian National University, Canberra, ACT 0200, Australia
[b] School of Electrical Engineering and Telecommunications, University of New South Wales, Sydney, NSW 2052, Australia

ARTICLE INFO

Article history:
Received 10 May 2009
Received in revised form 13 September 2009
Accepted 23 September 2009

Keywords:
Forensic voice comparison
Forensic speaker recognition
Forensic speaker identification
Paradigm shift
Likelihood ratio
Reliability

ABSTRACT

We are in the midst of a paradigm shift in the forensic comparison sciences. The new paradigm can be characterised as quantitative data-based implementation of the likelihood-ratio framework with quantitative evaluation of the reliability of results. The new paradigm was widely adopted for DNA profile comparison in the 1990s, and is gradually spreading to other branches of forensic science, including forensic voice comparison. The present paper first describes the new paradigm, then describes the history of its adoption for forensic voice comparison over approximately the last decade. The paradigm shift is incomplete and those working in the new paradigm still represent a minority within the forensic–voice-comparison community.

1. The new paradigm in forensic science

1.1. A paradigm shift

Today we are in the midst of what Saks and Koehler [1] have called a *paradigm shift* in the evaluation and presentation of evidence in the forensic sciences which deal with the comparison of the quantifiable properties of objects of known and questioned origin, e.g., DNA profiles, finger marks, hairs, fibres, glass fragments, tool marks, handwriting, and voice recordings. Saks and Koehler point out that they "use the notion of paradigm shift not as a literal application of Thomas Kuhn's concept [2], but as a metaphor highlighting the transformation involved in moving from a pre-science to an empirically grounded science" (p. 892). In Kuhnian terms, Saks and Koehler's paradigm shift might be better described as a shift from a pre-paradigm period towards a period where there is for the first time a single unifying paradigm for conducting normal science, i.e., a shift from a period during which a number of different schools pursue solutions to different sets of problems (with only partial overlap between sets) using different incompatible frameworks, towards a period during which there is agreement throughout the scientific community as to which problems are important (often a superset of the problems addressed by two or more of the pre-paradigm schools), and agreement as to the general procedures for solving these problems and the nature of suitable solutions. Whereas during the pre-paradigm period scientists must address a general audience and explain their theories from the beginning, during a normal-science period scientist principally address an audience which has already been educated in the fundamentals of the paradigm (e.g., by completing at least a bachelor of science degree) and they can immediately focus their efforts on a particular small question which forms part of the larger puzzle. Research efficiency and productivity is therefore greater during a normal-science period than during a pre-paradigm period.

Kuhn uses the term "paradigm" in two different senses, one broader and the other narrower: "On the one hand, it stands for the entire constellation of beliefs, values, techniques, and so on shared by the members of a given community. On the other, it denotes one sort of element in that constellation, the concrete puzzle–solutions which, employed as models or examples, can replace explicit rules as the basis for the solution of the remaining puzzles of normal science." [3] (p. 175). I will essentially be using the broader sense of "paradigm", which subsumes its narrower sense. Although I believe that Kuhn's thinking on scientific revolutions provides a useful tool for understanding the current situation in forensic science, and I point out a number of parallels below, one does not find a 100% correlation. One reason for this may be that forensic science is an applied science which must serve the imminent needs of society, and this consideration impinges to a greater extent than is the case in the natural sciences. In this, the forensic scientist is more like an engineer: "Unlike the engineer, and many doctors, and most theologians, the scientist need not choose problems because they urgently need solution and without regard for the tools available to solve them." [2] (p. 163).

Saks and Koehler [1] propose that a paradigm shift has already occurred in DNA profile comparison, and that other forensic comparison sciences are now shifting towards the new paradigm. In the present

☆ This is a revised version of an invited presentation given at the 2nd International Conference on Evidence Law and Forensic Science, Beijing, China, 25–26 July, 2009.

E-mail address: geoff.morrison@anu.edu.au.

1355-0306/$ – see front matter © 2009 Forensic Science Society. Published by Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.scijus.2009.09.002

paper my aim is to first describe the characteristics of the new paradigm, and then tell the story of its adoption to date in the field of forensic voice comparison.

1.2. The new paradigm

Saks and Koehler [1] describe the new paradigm as "empirically grounded science" (p. 892) as exemplified by "data-based, probabilistic assessment" (p. 893) as is current practice in forensic DNA comparison. They recommend that other forensic comparison sciences emulate DNA comparison, including that they "construct databases of sample characteristics and use these databases to support a probabilistic approach" (p. 893). They also make it clear that another important aspect of the new paradigm is the quantification and reporting of the limitations of forensic comparison via the measurement of error rates. The new paradigm therefore echoes the requirements for admissibility of scientific evidence set out in the US Supreme Court ruling in *Daubert v Merrell Dow Pharmaceuticals* (92–102) 509 US 579 [1993], which Saks and Koehler identify as a driving force for the paradigm shift. The Court ruled that, when considering the admissibility of scientific evidence, the judge must consider the methodology's scientific validity and evidentiary reliability, including whether it has been empirically tested and found to have an acceptable error rate. The call for other branches of forensic science to be more "scientific", emulate DNA profile comparison, and conform to the *Daubert* requirements was recently reiterated in the February 2009 release of the National Research Council (NRC) report to Congress on Strengthening Forensic Science in the United States [4]. Important aspects of a scientific approach identified in the report include "the careful and precise characterization of the scientific procedure, so that others can replicate and validate it; . . . the quantification of measurements . . .; the reporting of a measurement with an interval that has a high probability of containing the true value; . . . [and] the conducting of validation studies of the performance of a forensic procedure" (p. 4-8); the latter requiring the use of "quantifiable measures of the reliability and accuracy of forensic analyses" (p. S-16). The NRC report clearly recommends the use of more objective analytic methodologies over more subjective experience-based methodologies.

Although there does not appear to be any indication that either set of authors were consciously aware of this, there is one other component of the new paradigm which I believe is implicit in Saks and Koehler's [1] and the NRC report's [4] recommendation that other forensic comparison sciences emulate forensic DNA comparison: the adoption of the *likelihood-ratio framework* for the evaluation of evidence. In fact the NRC report consistently describes "identification" and "individualisation" as the aims of forensic science, which is antithetical to the use of the likelihood-ratio framework (see Section 1.4 below). The term "likelihood ratio" appears only once in the report, and this is in the title of a cited paper; however, the report recommends Aitken and Taroni [5], Evett [6], and Evett et al. [7] as providing "the essential building blocks for the proper assessment and communication of forensic findings"(p. 6-3), and all three advocate the use of the likelihood-ratio framework.

1.3. The likelihood-ratio framework

The leading *rôle* of forensic DNA comparison in the paradigm shift can in large part be attributed to the fact that it is a relatively new branch of forensic science which was put under extensive scrutiny when it was first presented in court in the late 1980s and early 1990s, and to the fact that it was developed by researchers who were trained and experienced in modern approaches to scientific research. The strong modern scientific background of those working in forensic DNA analysis arguably made it easier for them to understand and ultimately adopt what many forensic statisticians recommend as the logically correct framework for the evaluation of comparison evidence, the *likelihood-ratio framework*. Descriptions of the likelihood-ratio framework can be found in numerous textbooks and articles including Aitken and Taroni [5], Balding [8], Buckleton [9], Evett [10], Lucy [11], Robertson and Vignaux [12], and with specific reference to forensic voice comparison Champod and Meuwly [13], González-Rodríguez et al. [14], González-Rodríguez et al. [15], and Rose [16,17]. For a history of developments in forensic statistics prior to the advent of forensic DNA analysis (including use of the likelihood-ratio framework) see Evett [6], and for a history of statistical procedures applied to the evaluation of DNA evidence and the ultimate adoption of the likelihood-ratio framework in that field see Foreman et al. [18].

What follows is a brief description of the likelihood ratio framework. For simplicity, the description is provided only at the source level as this is the most relevant level for forensic voice comparison (see Cook et al. [19] on the hierarchy of *source*, *activity*, and *offence* propositions). The activity level is seldom important in forensic voice comparison because issues of transfer and persistence are seldom pertinent: voice recordings are usually deliberately recorded, and those presented for forensic analysis are typically associated with warrants and chain-of-custody documentation. Authentication of audio recordings, and analysis of disputed utterances, are normally considered to be areas of expertise which are distinct from forensic voice comparison. In forensic voice comparison one must, however, consider the effects of the conversion of the acoustic signal to an electronic signal and often its transmission over a telephone system, which can result in relatively poor quality voice recordings and potentially mismatches between the recording quality of known and questioned samples (transmission-channel effects). There may also be differences in speaking style, e.g., a lively telephone conversation on the recording of the questioned voice, and subdued answers to questions asked in a police interview on the recording of the known voice. The outcome of a forensic voice comparison may be of direct relevance for the offence propositions, for example, if the offence is uttering death threats and the questioned voice recording is a recording of someone uttering death threats.

In the likelihood-ratio framework the task of the forensic scientist is to provide the court with a *strength-of-evidence* statement in answer to the question:

> How much more likely are the observed differences/similarities between the known and questioned samples to arise under the hypothesis that they have the same origin than under the hypothesis that they have different origins?

The answer to this question is quantitatively expressed as a likelihood ratio, calculated using Eq. (1).

$$LR = \frac{p(E|H_{so})}{p(E|H_{do})} \tag{1}$$

where LR is the likelihood ratio, E is the evidence, i.e., the measured differences between the samples of known and questioned origin, H_{so} is the same-origin hypothesis, and H_{do} is the different-origin hypothesis. If the evidence is more likely to occur under the same-origin hypothesis than under the different-origin hypothesis then the value of the likelihood ratio will be greater than 1, and if the evidence is more likely to occur under the different-origin hypothesis than under the same-origin hypothesis then the value of the likelihood ratio will be less than 1. The size of the likelihood ratio is a numeric expression of the strength of the evidence with respect to the competing hypotheses. If the forensic scientist testifies that one would be 100 times more likely to observe the differences between the known and questioned samples under the same-origin hypothesis than under the different-origin hypothesis (LR = 100), then whatever