

CASE NOTE

Atkins v The Emperor: the ‘cautious’ use of unreliable ‘expert’ opinion

By Gary Edmond,^{*} Richard Kemp,[†] Glenn Porter,[‡] David Hamer,[§] Mike Burton,[¶] Katherine Biber^{**} and Mehera San Roque^{††}

Keywords Expert evidence; CCTV; Facial mapping; Admissibility; Reliability; Identification; Photograph

What happens to a country under constant surveillance? The recent decision in *Atkins v The Queen* provides a partial answer.¹ The sheer availability of images seems to be driving decisions about their admissibility and use as identification evidence. Confronted with CCTV recordings associated with criminal activities English courts have been reluctant to restrict their admission or impose limitations on the scope or form of incriminating opinion derived from them. Although the Court of Appeal decision in *Atkins v The Queen* is concerned primarily with the way in which an opinion derived from CCTV images was expressed, the decision exposes jurisprudential weakness and continuing problems with photo comparison and facial-mapping evidence.

^{*} Professor, Australian Research Council Future Fellow and Director, Program in Expertise, Evidence and Law, Centre for Interdisciplinary Studies of Law (CISL), School of Law, The University of New South Wales, Sydney 2052, Australia; email: g.edmond@unsw.edu.au. This work was supported by the Australian Research Council (DP0771770 and FT0992041).

[†] Senior Lecturer, Department of Psychology, The University of New South Wales. This research was supported by the Australian Research Council (DP0881623 and LX0883067).

[‡] Head, Forensic Sciences, University of Western Sydney.

[§] Associate Professor, School of Law, University of Sydney.

[¶] Professor, Department of Psychology, University of Glasgow.

^{**} Associate Professor, School of Law, University of Technology, Sydney.

^{††} Lecturer, School of Law, The University of New South Wales.

¹ *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876.

1. Ground of appeal: 'the issue of the expression of expert evidence'

In England, as in most other common law jurisdictions, opinion evidence is not generally admissible. 'Opinion based upon *identifiable expertise outside the experience of the jury* is one exception'.² According to the court in *Atkins* such 'expertise' must be 'identifiable' and:

*Courts need to be scrupulous to ensure that evidence proffered as expert, for any party, is indeed based upon specialised experience, knowledge or study. Mere self-certification, without demonstration of study, method and expertise, is by itself not sufficient.*³

At common law the judge is empowered to exclude opinion evidence where the witness is not appropriately qualified or experienced, or where the opinion is not based upon recognisable expertise. The appeal in *Atkins*:

... therefore does not raise any question as to the judge's power at common law to exclude evidence tendered as expert, if it be argued that the expert is *insufficiently qualified* or that his *evidence is insufficiently based upon expertise*.⁴

Where there is an objection to admissibility, it is for the party 'who tenders such evidence to establish the exception, viz the expertise'. The party that 'asserts admissibility must demonstrate it'.⁵ In *Atkins v The Queen* the admissibility of photographic comparison work by Mr Richard Neave was contested.⁶

Significantly, English courts have not consistently required 'knowledge' or 'reliability' before admitting incriminating opinions in serious criminal

² Ibid. at [9] (*italics added*).

³ Ibid. at [27] (*italics added*). The *Atkins* court also stated that: 'expert opinion evidence may be given by reference to studies which the expert has done but which a jury does not see' (at [17]).

⁴ Ibid. at [9] (*italics added*).

⁵ Ibid. at [9].

⁶ Richard Neave was the medical artist who prepared and presented the photographic comparison evidence. We do not mean to question Neave's expertise as a medical artist. The issue we raise is whether extensive experience as a medical artist—or as an anatomist (as in *R v Tang* [2006] NSWCCA 167) or in military intelligence—and experience in and out of courts provides an adequate basis for the provision of incriminating expert opinion evidence derived from images.

proceedings.⁷ In relation to facial mapping⁸—that is, the comparison of images for the purposes of helping to establish identity—cases such as *R v Stockwell*⁹ (also with Neave), *R v Clarke*,¹⁰ *R v Hookway*¹¹ (also with Neave), *R v Briddick*,¹² *Attorney-General's Reference (No. 2 of 2002)*¹³ and *R v Gardiner*¹⁴ support the proposition that ‘a suitably qualified expert with facial-mapping skills may properly give opinion evidence of his comparison of images from the scene, enhanced or otherwise, and a reasonably contemporary photograph of the accused’.¹⁵ Any dangers, such as unreliability, exaggeration and misrepresentation, are issues going to the weight assigned to the opinion at trial.

Because English courts have adopted a particularly liberal approach to the admissibility of incriminating expert opinion, including facial-mapping evidence,¹⁶ ‘Mr

7 Although the South Australian decision of *Bonython (R v Bonython (1984) 38 SASR 45)* has been influential and Courts of Appeal in *R v Luttrell* [2004] 2 Cr App R 31 and *R v Ciantar* [2005] EWCA Crim 3559 placed emphasis on the need for reliability as part of English admissibility jurisprudence. These are summarised in *Ciantar* at [21]: ‘The subject matter in the instant case, facial mapping, must be part of a body of knowledge or experience recognised to be a reliable body of knowledge. Secondly, the witness must be qualified to express an opinion by reason of his special acquaintance with that body of knowledge’. See T. Ward, ‘English Law’s Epistemology of Expert Testimony’ (2006) 33 *Journal of Law & Society* 572–95. Compare M. Redmayne, *Expert Evidence and Criminal Justice* (Oxford University Press: Oxford, 2001); A. Roberts, ‘Drawing on Expertise: Legal Decision Making and the Reception of Expert Evidence’ [2008] Crim LR 443–62.

8 It is not our intention to suggest that facial mapping (or body mapping) are recognisable ‘fields’. Even though English, and to some extent Australian, courts have recognised them as such, this recognition is largely limited to forensic contexts and seems misconceived (for some of the reasons described below). Similarly, the fact that some individuals have formed an organisation (see below n. 17) does not overcome fundamental epistemological problems. The emphasis on reliability in this note is intended to suggest that empirical evidence of ability and accuracy should be more important than legal recognition or professional ‘bootstrapping’.

9 (1993) 97 Cr App R 260.

10 [1995] 2 Cr App R 425.

11 [1999] Crim LR 750.

12 [2001] EWCA Crim 984.

13 [2003] 1 Cr App R 21.

14 [2004] EWCA Crim 1639.

15 *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [13] (*italics added*). In *Attorney-General's Reference (No. 2 of 2002)*, above n. 13 at [19], the court indicated that ‘where a witness who does not know the defendant spends substantial time viewing and analysing photographic images from the screen, thereby acquiring special knowledge which the jury does not have, he can give evidence of identification based on a comparison between those images and a reasonably contemporaneous photograph of the defendant, provided that the images and the photograph are available to the jury’. See also *R v Abnett* [2006] EWCA Crim 3320, [2006] All ER 244 at [14], [20].

16 Compare G. Edmond, ‘Pathological Science? Demonstrable Reliability and Expert Pathology Evidence’ in K. Roach (ed.), *Pediatric Forensic Pathology and the Justice System* (Queen’s Printer for Ontario: Toronto, 2008) 96.

Neave's expertise in his field was accepted' and the appeal in *Atkins* was limited to the use of the following graded scale (taken from an article by Bromby).¹⁷

Level	Description
0	Lends no support
1	Lends limited support
2	Lends moderate support
3	Lends support
4	Lends strong support
5	Lends powerful support

The question confronting the Court of Appeal in *Atkins* concerned Neave's 'expression of the level of support'—between “it lends support” and “lends strong support”—to the question of whether the accused (Dean Atkins) is the person in low quality incriminating CCTV images. There was, as Neave conceded, no database of facial features facilitating his move from noting apparent similarities between the images of Atkins and the perpetrator, to a conclusion about their significance. Nevertheless, the court explained:

... that where a photographic comparison expert gives evidence, properly based upon study and experience, of similarities and/or dissimilarities between a questioned photograph and a known person (including a defendant) the expert is not disabled either by authority or principle from *expressing his conclusion as to the significance of his findings*, and that he may do so by use of conventional expressions, arranged in a hierarchy, such as those used by the witness in this case and set out ... above.¹⁸

Such 'conclusions' are admissible provided their status as 'subjective opinion' is 'made crystal clear to the jury charged with evaluating them'.¹⁹ The absence of an 'expertise-based justification' and lack of a database does not prevent a *properly*

17 *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [9]. M. Bromby, 'At Face Value? The Use of Facial Mapping and CCTV Image Analysis for Identification' (February 2003) 153 *New Law Journal: Expert Witness Supplement* 302–4. The Forensic Imagery Analysis Group (FIAG) of the British Association for Human Identification (BAHID) has embraced this scale and produced a non-binding guidance sheet. This specious guide does not address the underlying image problems or the lack of a database.

18 *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [31] (*italics added*).

19 *Ibid.*

qualified expert from proffering an opinion about the identity of the person in incriminating images whether as *similarities* or in more *positive* terms.

While it might be thought that this approach is reasonable and that facial mappers should be entitled to express inculpatory opinions—and even though here we are primarily concerned with the Court of Appeal’s approach to ‘level of support’ scales—there are good reasons to doubt the reliability of much photographic comparison evidence. It should not be thought that fundamental epistemological problems can be overcome simply by recourse to ‘conventional scales’, putatively ‘cautious’ forms of practice or through the disclosure of problems—such as Neave’s many concessions. The failure to test the technique, or to have a clear idea about limitations and levels of error, means that many purported similarities are simply of unknown significance.

2. Like(rt) magic: expressions of confidence

Although the *Atkins* court excused the absence of a database, provided the absence is made ‘crystal clear’ to the jury, this concession does not allow us to move from apparent (or alleged) similarities to reliable opinions about identity. The significance of a database is that, were one available, it would facilitate this transition in a meaningful manner and do so with some understanding of the limitations of any conclusions.

(a) Similarities and differences: what a difference a database makes

At the heart of *Atkins* is the significance of *apparent* similarities and the *alleged* absence of differences between the images of Atkins and the perpetrator (see Fig. 1 below).²⁰ The question of whether we can be confident about similarities and differences logically precedes the question of what significance might be attached to actual similarities (and/or differences). The failure to produce a credible and reproducible technique for overcoming image distortions or to recognise the limitations imposed by image resolution (and other ‘risk factors’) means that we cannot be confident that alleged and apparent similarities (or differences) accurately represent the underlying reality.²¹

20 We use ‘apparent’ and ‘alleged’ because in the absence of non-contentious and highly discriminating ‘identifiers’ or a method for overcoming image distortions, similarities may be artefactual or epiphenomenal rather than real.

21 For a discussion of image problems see G. Edmond, K. Biber, R. Kemp and G. Porter, ‘Law’s Looking Glass: Expert Identification Evidence Derived from Photographic and Video Images’ (2009) 20 *Current Issues in Criminal Justice* 337–77.

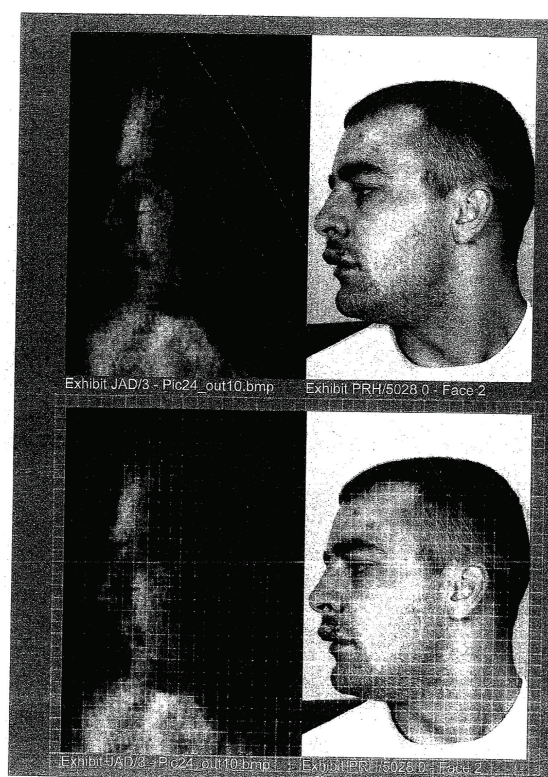


Figure 1. Copy of one of the set of images used by Neave. Consider the resolution of the enhanced crime scene image on the left and the addition of grid squares in the bottom frames. This copy image was taken from a PDF of the appeal submissions, and some loss of detail in the copy is likely. With permission, courtesy of Joe Stone.

Assuming that photographic interpretation was reliable (i.e. there was some demonstrated means of overcoming or eliminating the range of distortions and factors that affect reliability), what does it mean when we speak of similarities and differences?²² Here, the issue of databases of facial features and extrapolations from similarities assumes significance. To be useful, a database would include measurements or standardised descriptions of the features from each of a large number of faces sampled from several different population groups.²³ Such a

²² Here we are speaking about similarities and differences rather than the sometimes equally vexed question of interpreting what is happening: L. Schwartz, *Mechanical Witness: A History of Motion Picture Evidence in US Courts* (Oxford University Press, USA: 2009); N. Feigenson and C. Spiesel, *Law on Display* (NYU Press: New York, 2009).

²³ Though even this is deceptively complex. For there is no agreement about what should be measured and there is no, by way of example, standard working model of what constitutes a nose.

database would allow us to estimate how frequently a particular feature occurs in the relevant population and, critically, how often two or more features co-occur. This latter aspect is especially important. In order to derive any conclusions from the observation that images share more than one feature in common we must know about the degree of statistical independence of these features in the relevant population.²⁴ For example, it may be that a particular nose shape is often associated with a particular upper lip shape. In this case, the observation that both features occur tells us little more than the observation of either feature on its own. English (and Australian) facial mappers do not possess such a database and, without such or in the absence of testing, cannot confidently draw conclusions from patterns of similarities or differences in images.

To be clear, the court in *Atkins* is discussing the meaning of similarities in terms of the *identification* of Dean Atkins. That is, how does the person comparing images move from similarities to a level of confidence expressed as a 'level of support' for the proposition that two or more images are of the same person (or different persons—i.e. exclusion). Indeed, this is the question it formally addresses:

What is contended [by the appellant's barrister] is that it was impermissible for the expert to say that in his view those similarities lent something between support and strong support to the allegation that the man in the camera shot was Dean Atkins. ... It is contended that such an expert should never use any of these expressions but should be confined to identifying the similarities or dissimilarities between the faces compared.²⁵

And,

Mr Neave told the jury that he used these five different forms of expression. In the present case, he said that in his opinion the similarities which he had identified gave support to the proposition that the man on the camera was Dean Atkins which he put somewhere between 'the top of (3) and into (4).' This part of his evidence, as summarised by the judge, was in these terms:

'This comparison therefore offers a level of support for the allegation that Dean Atkins and the offender are one

²⁴ This is not unlike the problem with the opinion evidence presented by Sir Roy Meadow in *R v Clark* [2003] EWCA Crim 1020 and *R v Cannings* [2004] 1 All ER 725. It suggests the dangers of allowing even highly qualified experts to base opinions on long experience.

²⁵ *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [1].

and the same person, between “it lends support” and “lends strong support” to that conclusion. But you should remember this, that, as Mr Neave conceded, there is no database which would enable him to give a statistical analysis and so his scale is based on his own experience and expertise and, he added, in this case on the fact that he was able to exclude a large number of people, including one who was very similar to Dean Atkins, his brother Michael.²⁶

The lack of a database with which to interpret the significance of particular combinations of facial features is noted but *excused* and left to weight.²⁷ The absence of a database or some other credible method of assigning significance to purported similarities means the observer has no reasonable basis on which to draw conclusions about identity.²⁸ This makes it necessary to invent a value (like ‘lends strong support’), which is simply a subjective summation (or impression) lacking methodological rigour.

Rather than focus on limitations inherent in the process by which the observer moves from alleged similarities to identification, the effect that apparent similarities and intuitive interpretations might have on the accuracy of the witness’s conclusion, or the decision-making of lay fact-finders, courts simply require the absence of a database—and not necessarily the significance or implications of its absence—to be made transparent.

... the absence of a database is in particular a matter which must be explained unequivocally to any jury considering such evidence.²⁹

And,

On principle, we accept the caution with which any expression of conclusion in relation to evidence of this kind (and others) needs to be approached. We agree that the fact that a conclusion is not based upon a statistical database recording the incidence of the features compared as they appear in the population at large *needs to be made crystal clear to the jury*. But we do not agree that the absence of such a database means that no opinion can be expressed by the witness beyond rehearsing his

²⁶ Ibid. at [8]–[9].

²⁷ See also *R v Clarke* [1995] 2 Cr App R 425 at 431–2.

²⁸ Cf. *Davie v Lord Provost, Magistrates and Counsellors of the City of Edinburgh* (1953) SC 34 at 39–40.

²⁹ *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [16].

examination of the photographs. *An expert who spends years studying this kind of comparison can properly form a judgment as to the significance of what he has found in any particular case.*³⁰

This means that where the witness concedes that there is no database—so that their opinion is not probabilistic in any empirical sense—they are nevertheless allowed to present a subjective incriminating opinion about the significance of purported similarities because of their extensive experience:

... as Mr Neave conceded, there is no database which would enable him to give a statistical analysis and so his scale is based on his own experience and expertise ...³¹

(b) Problems with recourse to the Bromby scale: detailing the Emperor's clothes

In *Atkins* we are told that the images were partial, fleeting and of poor quality. Nevertheless, the witness concluded that:

... he could find no difference between the photograph of Dean Atkins and the CCTV image on morphological examination and identified particular similarities in eight different areas, of which the principal were nose, mouth, upper lip and chin. The camera angles from which the photograph and the CCTV image were taken were *virtually identical and thus sufficient to enable grid measurements to be made*. These were also *consistent with* the photograph of Dean, and so was the product of the two superimposition or substitution tests.³²

In the absence of a database *and* a set of procedures or standards for interpreting photographs, we can only speculate about how photo comparison witnesses transform apparent similarities (and lack of differences) into a level of confidence about whether the person of interest is the suspect/accused.

Given the lack of a suitable database, the problem is on what basis can a witness move from the interpretation of apparent similarities between person(s) in two (or more) images to making claims about the meaning of those apparent similarities. Actually, this predicament is neatly summarised by counsel for Atkins:

30 Ibid. at [23] (italics added).

31 Ibid. at [8].

32 Ibid. at [6] (italics added). Use of the phrase 'consistent with' is critically analysed in several Canadian inquiries into wrongful convictions. See below n. 63.

... the absence of any database cataloguing the numbers of persons with particular facial features or measurements meant that there was no expertise-based justification for his expression of the level of support which that comparison gave to the allegation that the man in the photograph was Dean Atkins.³³

As cursory inspection suggests, and the Court of Appeal confirms, the images available for analysis in *Atkins* were of low resolution and 'far from clear to the naked eye'.³⁴ Nevertheless, the facial mapper was willing—on the basis of a few apparent similarities (of unknown frequency or independence), an apparent inability to discern any dissimilarities, and *practical indifference* to the many 'risk factors' (some of which were conceded)—to testify that there were only similarities between the incriminating image and a reference image of Dean Atkins. Neave concluded that the incriminating image 'lent support' or 'strong support' to the contention that it was Dean Atkins at the crime scene. Here there is an implicitly unproblematic move from limited similarity to a conclusion about identity (or individualisation). In the absence of information about the frequency with which particular facial characteristics occur in the population, or how commonly these features exist in combination, the witness was prepared to say that one or more of the very low quality images provided 'support' or 'strong support' to his conclusions about identity.

Not only is there a conceptual problem in moving from one or more apparent similarities to identity, there is also the serious question of what kind of evidence might properly ground support, let alone 'strong' and 'powerful' support.³⁵ The absence of an articulated and validated method, the inability to address the frequency of facial features, or even address, let alone overcome, the reliability issues suggest that the level of confidence apparent from the witness's use of the Bromby scale is not merely misguided, but misleading.

33 Ibid. at [10]. Some of these difficulties were recognised by Mitting J in *R v Gray* [2003] EWCA Crim 1001. Although that decision is marginalised by the court in *Atkins* (at [12]–[16]) on the basis that it is unreported, the particular facial-mapping expert 'overstepped the mark' and was shown in two other cases to have been wrong. Rather than invoke the 'bad apple' theory of experts, courts should pay serious attention to the reliability of methods and conclusions. For, in the absence of evidence of trustworthiness, how can we be confident that currently acceptable comparison experts will not also be exposed as mistaken?

34 *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [6]. The authors have seen an electronic copy of the images featured in the appeal documents. See Fig. 1 above.

35 The issues are not unique to photographic comparison evidence: see e.g. S. Cole, 'Is Fingerprint Identification Valid? Rhetorics of Reliability in Fingerprint Proponents' Discourse' (2006) 28 *Law & Policy* 109–35.

We should not quibble over semantics or underestimate the significance and potential unfair prejudice of incriminating opinion evidence in the adversarial criminal trial. In the context of serious criminal proceedings a person with long experience as a medical artist, who had previously been allowed to express opinions about similarities and identification, was qualified as an expert and allowed to express the opinion that images of a person from a crime scene lent support to the Crown contention that Dean Atkins was the perpetrator.³⁶ Regardless of any claims to the contrary, this evidence was presented and likely understood as positive identification evidence whereas, in reality, all the witness's analysis tells us is that his interpretation of the shapes drawn from the images of the perpetrator does not match the appearance of Atkins' brother Michael, their co-accused Carty, and a small number of known burglars living in the vicinity of Uxbridge.³⁷ Neave cannot know, and so does not tell us, how many other individuals are likely to possess the apparent features noted in the images. It may be that many men in the United Kingdom would match this image, or it could be that no other person does so, but the jury were told that only Dean Atkins could not be excluded. Neave might not have positively identified the accused but he could see only similarities between Dean Atkins and the person of interest. To put this another way, he could discern no differences and the photographic evidence, according to this experienced practitioner, 'lent support', at the very least.

Facial mappers may choose to employ the Bromby scale in an attempt to make their opinion appear more *scientific*.³⁸ However, recourse to this scale makes the absence of an underlying method more conspicuous. In some ways the illusory precision of the scale creates problems for the witness. Had Neave instead expressed his own intuitive impression, in his own words, this might have been even more difficult to contest. Unwittingly, recourse to the Bromby scale serves to reinforce the very serious methodological flaws and limitations with photographic comparison evidence and expertise.

In some ways Neave's description of the process by which he superimposed grid squares on the various images parallels the illusion created by the Bromby scale.

36 Identity is, in effect, the ultimate issue. While the ultimate issue rule is now practically moribund, courts should nevertheless be careful when those whose expertise is questioned endeavour to express opinions on matters such as identity. See e.g. *R v GK* [2001] NSWCCA 413 at [40].

37 Without more, the fact that these features are not similar to (or shared by) other burglars of Uxbridge is trivial though presumably misleading and unfairly prejudicial. The presentation of these individuals as 'foils', especially where the analyst is not 'blinded', is also misleading and adds little if anything to the reliability of any photographic comparison exercise.

38 There is an extensive literature suggesting that people cannot (or, at least, do not) use such scales consistently. See e.g. L. Moxey and A. Sanford, *Communicating Quantities: A Psychological Perspective* (Erlbaum: Hove, 1993).

Use of the grid implies a level of precision (whether anthropometric or morphological) that is also without foundation. In the absence of a credible means of overcoming photographic distortion and other forms of unreliability, how can the witness be confident that the features in each of the grid squares accurately represent what the person of interest actually looks like? Can we, for example, meaningfully compare the forms in the grid squares—whether individually or in combination? Once again, a seemingly scientific apparatus, the grid, lends the appearance of precision.³⁹

In the end the inability to overcome distortion *in conjunction with* the lack of a database means that all the witness can say is that in his opinion parts of the mouth, nose and chin *appear similar* in these partial, poorly resolved images. The absence of information about the frequency of facial features individually or in combination means that the witness cannot take this purported 'match' evidence any further.

To draw an analogy, imagine that a biologist presented DNA evidence along the lines that her analysis 'lent very strong support' to the conclusion that a badly degraded trace sample yielding a partial profile came from the defendant because it showed several similarities to a reference sample taken from the defendant and she was unable to discern any differences between the samples. Now imagine that under cross-examination this biologist conceded that she knew nothing about the extent of the degradation of the sample and how its quality affected the reliability of her conclusions, had no knowledge of the frequency of the features in the partial profiles (apart from what she had encountered in her laboratory work), and that there was no data in existence which allowed her to estimate the frequency with which the partial profile appeared in the population.⁴⁰ It is highly unlikely that such evidence would be presented to a jury, even with the proviso that the jury be told about the lack of a database and the opinion being experience-based. Why then do courts choose to admit opinion evidence of a similar quality when it concerns visual identification from images?

Perhaps the difference lies in our relative experience with biological trace evidence and faces. We are all, to some degree, experts in face recognition (and experienced with photographs and film via engagement with media and advertising). However, familiarity with faces may encourage *experts* and judges (and jurors) to

39 See e.g. T. Porter, *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life* (Princeton University Press: Princeton NJ, 1995).

40 Compare J. Aronson, *Genetic Witness: Science, Law, and Controversy in the Making of DNA Profiling* (Rutgers University Press: Piscataway NJ, 2007); M. Lynch, S. A. Cole, R. McNally and K. Jordan, *Truth Machine: The Contentious History of DNA Fingerprinting* (University of Chicago Press: Chicago, 2008).

underestimate the difficulties involved in facial and photographic comparisons.⁴¹ Because most of us know relatively little about DNA, it is easy to see that comparison is complicated and that in order to make meaningful probabilistic conclusions regarding identity, we must employ standardised techniques and make reference to databases (and assumptions about populations).⁴² However, empirical research indicates that facial comparison is also complicated and, when it comes to strangers, surprisingly inaccurate.⁴³ Moreover, empirical research suggests that extraneous cues, such as circumstantial evidence or the opinions of an *expert*, are likely to dramatically affect assessments regardless of their reliability.⁴⁴

3. Burden shifting: substituting concessions and excuses for validity and reliability

Rather than require the Crown actually to obtain opinions using demonstrably reliable methods and repeatable techniques, trial and appellate courts have exhibited a willingness to allow the incriminating opinions of investigators to go before the jury on condition that some limitations—such as general problems with image reliability or the lack of a database—are disclosed. This obliges the defence to negate the speculative, though purportedly cautious, opinions of experts rather than require the experts to demonstrate to the court that they can actually do

41 The idea that photographs produce reliable evidence or a sense of truth through their *objective* rendering may also be a reason why many images are allowed to go before juries. Compare J. Mnookin, 'The Image of Truth' (1998) 10 *Yale Journal of Law & the Humanities* 1–74; C. Guilshan, 'A Picture Is Worth a Thousand Lies: Electronic Imaging and the Future of the Admissibility of Photographic Evidence' (1992) 18 *Rutgers Computer & Technology Law Journal* 365–80; M. Sturken and L. Cartwright, *Practices of Looking: An Introduction to Visual Culture* (Oxford University Press: 2001); G. Porter, 'Visual Culture in Forensic Science' (2007) 39 *Australian Journal of Forensic Sciences* 81–91.

42 This has not prevented mistakes and numerous appeals over procedures and the expression of results.

43 Humans are very good at recognising familiar faces, even from poor quality images: however, we appear to be unaware that our expertise is limited to the faces of those we know well, and as a result we overestimate our ability to determine whether two images are of the same unfamiliar person (i.e. unfamiliar face matching). See J. Davis and T. Valentine, 'CCTV on Trial: Matching Video Images with the Defendant in the Dock' (2008) 23 *Applied Cognitive Psychology* 482–505; M. Burton *et al.*, 'Face Recognition in Poor-quality Video: Evidence from Security Surveillance' (1999) 10 *Psychological Science* 243. This overconfidence also affects other fields where identification is important. For example, governments rely on photographic identification documents such as passports to establish the identity of people crossing borders. However, psychological research shows that even people charged with checking these documents, and trained to do so, are likely to make many errors. See R. Kemp, N. Towell and G. Pike, 'When Seeing Should Not Be Believing: Photographs, credit cards and fraud' (1997) 11 *Applied Cognitive Psychology* 211–22.

44 For example, in a laboratory study undertaken by one of the authors, participants were asked to decide if a CCTV image featured the 'defendant'. Their decision was better predicted by the *expert* opinion evidence they received on the matter than by the truth of the situation (i.e. whether the 'defendant' was actually in the images): R. Kemp, S. Heidecker and N. Johnston, 'Identification of Suspects from Video: Facial Mapping Experts and the Impact of their Evidence', paper presented at the 18th Conference of the European Association of Psychology and Law, Maastricht, 2–5 July 2008.

what they claim. The decision in *Atkins* exemplifies the substitution of concessions and excuses for validation studies and other indicia of reliability.⁴⁵

In *Atkins* the witness makes numerous concessions. These include:

In evidence before the jury Mr Neave identified nine different factors which could affect the reliability of his exercise. Many related to the quality of the photographs and the various things which could affect that. Another was pixilation, which is or can be image compression when a photograph is transferred to a computer. Two more of his stated *risk factors* were:

- i) the possibility that two different people may appear indistinguishable; and
- ii) the fact that there exists no database of facial characteristics.

He also dealt with the particular quality of the CCTV photograph of the offender, which was significantly limited by short duration, some element of distortion at one edge, shadow and the sudden change from dark to light when the door was opened and the person appeared with light to one side beyond him.⁴⁶

An observer might wonder at what point the cumulative effect of these concessions (or 'risk factors') actually prevents photo comparison witnesses from proffering expert opinions. In the absence of serious testing, published studies, documented means of overcoming image problems, attempts to develop databases or other means of assuring accuracy, allowing photo comparison witnesses to testify is not unlike admiring the performance of the *weavers* in Hans Christian Andersen's *The Emperor's New Clothes*. As it turns out, the mere concession of problems, like the 'self-certified' ability of the 'weavers', is not enough.⁴⁷

⁴⁵ This is not unlike 'the peculiar property of being invisible to any person who was unfit for his position or inexcusably stupid' in Andersen's 'The Emperor's New Clothes' in *Fairy Tales* (2004, originally published 1837) 91.

⁴⁶ *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [7] (italics added).

⁴⁷ See C. Wilkinson and R. Evans, 'Are Facial Image Analysis Experts Any Better than the General Public at Identifying Individuals from CCTV Images?' (2009) 49 *Science & Justice* 191, 195. This paper misleadingly suggests, on the basis of a limited study, that because the authors scored better than a selection of lay people 'the results ... support the utilisation of expert evidence in court and these experts have demonstrated skills, knowledge and abilities superior to the general public'. However, the real problem is, not whether so-called experts perform better than average, but whether their opinions are sufficiently reliable to be admitted as evidence and whether criminal trials can or should be obliged to cope with apparently error-prone opinions. Relative abilities should not be conflated with reliability.

Concession does not inoculate against unreliability (or the very real risk that the jury might misuse the opinion).⁴⁸ Ordinarily, conceding serious and persistent problems will provide a basis for the exclusion of evidence rather than a justification for admission.

In addition, it is important to stress that the ‘technique’ used by the witness in *Atkins* is not a standardised process but rather an incoherent amalgam of axioms, commitments, descriptions and practices. Moreover, claims about the ability to exclude, like claims made for superimposition and substitution ‘tests’, are misleading.⁴⁹ The inability to explain how ‘risk factors’ are credibly managed means that additional ‘tests’ merely compound the existing limitations or inspire inappropriate confidence rather than provide additional or independent checks. Many of the ‘techniques’, whether dependent on quantification or qualification, assume that image distortion is overcome—without explaining how—and that the observer can make sense of resulting similarities and differences experientially.

A few additional points are worth making. First, the various concessions (in *Atkins* and elsewhere) appear to have emerged largely in response to cross-examination and criticism by defence experts (and others) rather than through study or experimentation. Few seem to have been volunteered by facial mappers. Secondly, making concessions does not overcome the problems inherent in the concessions. If there are image reliability problems and there is no method for overcoming them, then why should we be interested in apparent similarities between images? Similarly, while it is important to acknowledge the lack of a database, acknowledging its absence does not provide a credible basis for allowing the witness to express an opinion that ought to be derived from a database or some other empirical foundation. Thirdly, Neave contends that the camera angles and perspectives used to obtain the reference photos were ‘virtually identical’.⁵⁰ Just how similar they need to be is not explained or disclosed. This is one of the problems with the ‘technique’. How does the witness know they are taken from the same angle and with the same lens? Is it supported by the appearance of similarities in the various images and therefore tautological? How different do the angles, lenses and lighting have to be before the exercise is invalid or unreliable? Even a cursory examination of the reference images used by the analyst reveals that the photographs of the Atkins brothers and their co-accused, Carty, are not identical in circumstance or orientation. In the reference photograph of Dean Atkins, supposedly a facial profile, both eyebrows can clearly be seen. The

48 On ‘inoculation’, see R. Barthes, *Mythologies* (Paladin: London, 1973).

49 *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [5], [6].

50 *Ibid.* at [6].

reference photographs of Carty and Michael Atkins are taken from a side perspective whereas in the reference photograph of Dean Atkins the neck is turned away to simulate that appearance. Do these differences matter and if not why not? How much variation can be tolerated before the comparisons lose any reliability? Fourthly, it is common practice for those engaged in photographic comparison to be told who the police believe the perpetrator is and to be given only two sets of images to compare. In the absence of 'blinding', experts are likely to be biased by this information. There is compelling evidence that even experts from established fields with quite elaborate formalised methods and training encounter difficulty in overcoming such influences.⁵¹ Fifthly, the witness's 'technique', axioms and approach are rhetorically based on some of the approaches now routinely applied to DNA evidence. However, such parallels have limitations. For photographic evidence might in some circumstances lead to a positive identification (especially when familiars are involved or discriminating features such as the appearance of very unusual scars, blemishes or tattoos are not in dispute).⁵²

In practice, substituting concessions (or declarations of inabilities) for evidence of validity and reliability effectively imposes a reverse onus upon the defence. Rather than require the Crown to demonstrate that incriminating expert opinion evidence has probative value, incriminating opinions are admitted and left to the defence to negate. Rather than exclude unreliable expert opinions and expert opinions of unknown probative value, English judges leave incriminating opinion to the trial and the fact-finder. They place great faith in a range of formal protections (such as defence lawyers, cross-examination, opposing experts and judicial directions), the lay jury and appellate courts.⁵³ Admissibility jurisprudence and discretionary exclusions are rarely invoked to exclude incriminating

51 In Australian cases the letter of instruction prepared by investigators frequently includes only images of the perpetrator and the prime suspect. The implication, not always unstated, is that the images are of the same person. Normally no alternative suspects are provided and in at least one case the letter of instruction noted that the suspect had changed appearance, but invited the analyst to use his experience to see through this attempted subterfuge. This is clearly a very suggestive procedure and empirical evidence indicates that suggestions more subtle than these can have profound effects on the decision-making of experts and non-experts alike. See I. Dror, D. Charlton and A. Peron, 'Contextual Information Renders Experts Vulnerable to Making Erroneous Identifications' (2006) 156 *Forensic Science International* 74–8; I. Dror and R. Rosenthal, 'Meta-analytically Quantifying the Reliability and Biasability of Forensic Experts' (2008) 53 *Journal of Forensic Sciences* 900–3.

52 These are very rarely, however, unique identifiers; even though they might be highly probative (once image reliability issues are overcome or circumvented).

53 These are increasingly open to a range of critical perspectives, especially in the United States. See K. Findley, 'Innocents at Risk: Adversary Imbalance, Forensic Science, and the Search for the Truth' (2008) 38 *Seton Hall Law Review* 893–974; B. Garrett and P. Neufeld, 'Invalid Forensic Science Testimony and Wrongful Convictions' (2009) 95 *Virginia Law Review* 1–97.

opinion evidence of unknown probative value notwithstanding the dangers of unfair prejudice to the accused.⁵⁴ In consequence, the accused is confronted with incriminating opinion of unknown reliability in the context of a circumstantial case. The incriminating opinions are admissible on the condition that defects, arguably fundamental and irreparable defects, are disclosed. This makes no sense with regard to opinion evidence used routinely in the investigation and prosecution of crime, especially where it could be tested for validity and reliability.

Appellate courts defend this stance on the basis that there are many areas of forensic science where there is no database and other experts are routinely permitted to express opinions.

That conclusion is consistent with the practice of experts in a number of fields to use this kind of scale of expressions of opinion. Such expressions are commonly encountered, as counsel on both sides confirmed and our own experience suggests, in many other fields of comparison. Some random examples are comparisons of handwriting or fibres, glass fragments or footwear patterns. In some of those there is some, often very incomplete, data about the incidence of general occurrence; in others there is no database.⁵⁵

However, the fact that courts have traditionally allowed opinion evidence of unknown, and frequently questionable, epistemic provenance should not be used as a basis for the continued admission of incriminating opinion evidence. On this point it is useful to contrast English *laissez-faire* with recent critical work emanating from the United States and Canada.⁵⁶

⁵⁴ G. Edmond, 'Specialised Knowledge, the Exclusionary Discretions and Reliability: Re-assessing Incriminating Opinion Evidence' (2008) 31 *University of New South Wales Law Journal* 1–55.

⁵⁵ *Dean Atkins and Michael Atkins v The Queen* [2009] EWCA Crim 1876 at [26]. Interestingly, many of these other disciplines have a range of formal protocols designed to avoid some of the problems encountered in facial mapping. Most forms of photographic comparison—such as footwear impressions, by way of example—require very high quality images free from a range of distortions. This stands in stark contrast with the images used in *Atkins*.

⁵⁶ National Research Council, *Strengthening Forensic Science in the US: A Path Forward* (Academies Press: Washington DC, 2009); S. Goudge, Commissioner, *Report of the Inquiry into Pediatric Forensic Pathology in Ontario*, 1 October 2008, available at <<http://www.goudgeinquiry.ca/>>, accessed 4 February 2010; M. Saks and D. Faigman, 'Failed Forensics: How Forensic Science Lost Its Way and How It Might Yet Find It' (2008) 4 *Annual Review of Law & Social Science* 149; M. Saks and J. Koehler, 'The Coming Paradigm Shift in Forensic Identification Science' (2005) 309 *Science* 892–5. Even the UK Law Commission's recent Consultation Paper places a premium on reliability, though its receptiveness to opinions based on 'expertise' and 'experience' threatens to admit the kinds of evidence that a reliability threshold might exclude: Law Commission, *The Admissibility of Expert Evidence in Criminal Proceedings in England and Wales: A New Approach to the Determination of Evidentiary Reliability*, Consultation Paper No. 190 (2009).

Finally, the existence of other incriminating evidence—like the provision of similarity evidence on previous occasions—should not be used to support the admission of incriminating expert opinion evidence. Incriminating expert opinions need to be assessed on their own terms.⁵⁷ Either the expert can do what he or she claims, and can provide evidence to support that contention, or they cannot. The absence of evidence necessarily raises questions about expertise.⁵⁸ Long experience is not enough where there is a reasonable expectation that the witnesses would have actually undertaken studies, published results or tested themselves.⁵⁹ Almost two decades after (Neave appeared in) *Stockwell*, it is not enough to just identify problems about which to warn juries. Rather, courts, prosecutors, investigators and the institutionalised forensic sciences must be more proactive. It would be relatively straightforward for facial mappers (and specialist photo comparison witnesses) to conduct proper tests of the reliability of their evidence, and such tests should be seen as a minimum requirement for the presentation of this evidence.⁶⁰ Similarly, it is not unreasonable to expect that over the last 20 years those working with images should have drawn up a set of standard operating procedures designed to minimise error and should be in a position to provide some information about the confidence intervals surrounding their conclusions.⁶¹ Courts would then be in a position to determine *what level of reliability is acceptable* and how to present the evidence to jurors rather than excusing incriminating opinions because the state and the experts are unwilling to undertake the necessary research.

In the meantime, if it is not possible to describe procedures for comparing images that are demonstrably reliable, then that would seem to provide a proper basis for the exclusion of expert photographic comparison evidence. The premium placed on liberty, fairness and factual rectitude means that the accusatorial criminal

57 See e.g. *R v Trochym* [2007] 1 SCR 239 at [31]–[32].

58 H. Collins and R. Evans, *Re-thinking Expertise* (University of Chicago Press: Chicago, 2007).

59 Similarly, claims about impartiality, professional duties and standards and even novel methods of eliciting expert opinion in the courtroom do not overcome the failure to ascertain validity and reliability. See e.g. G. Edmond, 'Merton and the Hot Tub: Scientific Conventions and Expert Evidence in Australian Civil Procedure' (2009) 72 *Law & Contemporary Problems* 159.

60 See K. Kleinberg, P. Vanezis and M. Burton, 'Failure of Anthropometry as a Facial Identification Technique Using High-quality Photographs' (2007) 52 *Journal of Forensic Science* 779.

61 The Chief Constables (UK) developed a loose set of recommendations for photographic comparison, but these are so vague as to have very little impact on practice, admissibility or reliability. See Association of Chief Police Officers of England, Wales and Northern Ireland, *National Working Practices in Facial Mapping* (2003), especially Appendix 3.

trial is not the place to allow investigators and those with long experience or impressive credentials to express their incriminating opinions.⁶²

4. Conclusion: more reliable suits?

The *Atkins* appeal addresses the way a photo comparison *expert* might express a subjective opinion. Curiously, earlier courts, and this is upheld in *Atkins*, suggest that such a witness might, notwithstanding image distortion or the non-existence of a database (and notwithstanding the apparent unwillingness of this particular witness to individualise), give positive identification evidence. *Atkins*, in effect, represents a jurisprudential backwater, largely indifferent to scientific processes, the reliability of incriminating opinions, or the way fact-finders might understand such evidence.

Recently, in Canada and the United States, judges, lawyers, psychologists and forensic scientists have been wrestling with how to express incriminating opinions. This has been a vexed issue, linked closely to several wrongful convictions.⁶³ In response to North American experience and the appeal in *Atkins* it is important to recognise that there is *no established method* for attaching evidentiary significance to alleged or apparent similarities. Even ignoring the fundamental issues of image distortion and interpretation, there are problems associated with how often different individuals might appear ‘indistinguishable’ and how common particular facial features or combinations of facial features are among relevant sub-populations—especially within families. Consequently, as things stand there is no reasonable basis for allowing a person, however skilled in anatomy, IT, photography, military surveillance, psychology or medical art, to extrapolate about the significance of apparent similarities. In addition, there may

62 Arguably, in the absence of demonstrably reliable techniques, photographic comparison evidence should only be used for the purposes of exclusion. Such an approach is not based upon the ability of facial mappers to identify accurately differences (as opposed to similarities), but rather reflects the way a criminal justice system committed to avoiding the conviction of innocent persons might manage the risk. Contrast the slightly duplicitous symmetry implied in *Atkins* at [25]. On the use of investigators as experts, see G. Edmond and M. San Roque, ‘Quasi-Justice: Ad Hoc Experts and Identification Evidence’ (2009) 32 *Criminal Law Journal* 8.

63 Goudge, *Report of the Inquiry into Pediatric Forensic Pathology in Ontario*, above n. 56; P. LeSage, Commissioner, *Report of the Commission of Inquiry into Certain Aspects of the Trial and Conviction of James Driskell*, January 2007, 149, 172, available at <<http://www.driskellinquiry.ca/>>; *Report of the Kaufman Commission on Proceedings Involving Guy Paul Morin*, 1998, 340–1, available at <<http://www.attorneygeneral.jus.gov.on.ca/english/about/pubs/morin/>> (reports accessed 4 February 2010); J. Dwyer, P. Neufeld and B. Scheck, *Actual Innocence* (Random House: 2000). See also D. McQuiston-Surrett and M. Saks, ‘The Testimony of Forensic Identification Science: What Expert Witnesses Say and What Fact Finders Hear’ (2009) 33 *Law & Human Behavior* 436; A. Sanford and L. Moxey, ‘New Perspectives on the Expression of Quantity’ (2003) 12 *Current Directions in Psychological Science* 240.

even be problems in leaving *incriminating* images before lay juries. We cannot simply assume that fact-finders are capable of overcoming some of the illusions created by image distortion, particularly in the context of a criminal trial where photographic interpretations are likely to be influenced—consciously and unconsciously (and perhaps irrationally)—by supplementary information.

While courts are obviously free to disregard mainstream approaches to scientific research and expertise, they do so at considerable peril. Moreover, where they do disregard orthodox methodological approaches they should have good reasons—preferably good institutional reasons consistent with legal principle and institutional values. In the accusatorial trial it is becoming increasingly difficult to understand the continued admission of photographic comparison evidence given: the state carrying the burden of proof; growing concern that defence mechanisms do not consistently identify, expose or convey problems with expert opinion; empirical concerns about jury comprehension and image interpretation; along with the ability actually to assess the proficiency of photo comparison analysts. It might be timely for senior judges (or legislators) to begin to review their approaches to incriminating expert opinion in ways that embody the values of an accusatorial system and the best advice of scientists (and empirical researchers).⁶⁴

In the absence of any indication of the reliability or validity, and given the concerns we have raised about the unjustified assumptions underlying this evidence, it is appropriate to be sceptical about its probative value. There are good reasons to exclude this evidence on the grounds of unwillingness or inability to demonstrate validity and reliability, inability to overcome (as opposed to recognising and conceding) problems, the speculative nature of the expressions, and the fact that witnesses tend to be biased by the way they are introduced to the images and undertake their comparisons. The issue of attaching significance to *similarities* raised in *Atkins* is a problem, but it is just one of a myriad of serious and unaddressed problems that have become part of the way police, lawyers and courts in England (and elsewhere) approach photographic comparison evidence.

⁶⁴ House of Commons, Science and Technology Committee, *Forensic Science on Trial* (2005); National Research Council, *Strengthening Forensic Science in the US: A Path Forward* (National Academies Press: Washington DC, 2009); E. Beecher-Monas, *Evaluating Scientific Evidence* (Cambridge University Press: 2006). See also G. Edmond and D. Hamer, 'Bacon's Chickens: Empirical Studies of Evidence Law' in P. Cane and H. Kritzer (eds.), *The Oxford Handbook of Empirical Legal Studies* (2010) (forthcoming).

CASE NOTE

‘But he doesn’t have anything on!’ said a little child. ‘Good Lord, listen to the voice of the innocent,’ said the father. And one person whispered to the next what the child had said. ‘But he doesn’t have anything on!’ the entire crowd cried at last. The Emperor cringed, because he thought they were right, but then he reasoned, ‘I have to make it through the procession.’ He held himself even prouder than before, and the chamberlains walked along carrying the train that was not there at all.⁶⁵

⁶⁵ Hans Christian Andersen, ‘The Emperor’s New Clothes’ in *Fairy Tales Told for Children: First Collection* (1837) 94–5.