

1999

DNA Database Searches and the Legal Consumption of Scientific Evidence

Richard D. Friedman


University of Michigan Law School, rdfrdman@umich.edu

Peter Donnelly

Wellcome Trust Centre for Human Genetics, directorpa@well.ox.ac.uk

Available at: <https://repository.law.umich.edu/articles/942>

Follow this and additional works at: <https://repository.law.umich.edu/articles>

 Part of the [Courts Commons](#), [Evidence Commons](#), [Science and Technology Law Commons](#), and the [Supreme Court of the United States Commons](#)

Recommended Citation

Friedman, Richard D. "DNA Database Searches and the Legal Consumption of Scientific Evidence." P. Donnelly, co-author. *Mich. L. Rev.* 97, no. 4 (1999): 931-84.

This Article is brought to you for free and open access by the Faculty Scholarship at University of Michigan Law School Scholarship Repository. It has been accepted for inclusion in Articles by an authorized administrator of University of Michigan Law School Scholarship Repository. For more information, please contact mlaw.repository@umich.edu.

DNA DATABASE SEARCHES AND THE LEGAL CONSUMPTION OF SCIENTIFIC EVIDENCE

*Peter Donnelly**
and
*Richard D. Friedman***

DNA evidence has transformed the proof of identity in criminal litigation, but it has also introduced daunting problems of statistical analysis into the process. In this Article, we analyze a problem related to DNA evidence that is likely to be of great and increasing significance in the near future. This is the problem of whether, and how, to present evidence that the suspect has been identified through a DNA database search. In our view, the two well-known reports on DNA evidence issued by the National Research Council ("NRC"), each of which has carried great authority with the American courts on various issues, have been badly mistaken in their analysis of this problem. Similar errors affect the analyses of scholars who have supported the NRC reports. We will also offer some reflections on the habits of mind, of both lawyers and statisticians, that may have led to this result. Finally, we will suggest an approach that legal decisionmakers might take in general with respect to scientific and statistically-based evidence to avoid this kind of difficulty. This approach is significantly different from that which pervades the Supreme Court's decisions in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*¹ and its recent sequel, *General Electric Co. v. Joiner*,² as well as their precursor, *Frye v. United States*.³ And the

* Professor of Statistical Science and Head of the Department of Statistics, University of Oxford. B.Sc. [Hons.] 1980, Queensland; D.Phil. 1983, Oxford. — Ed. Part of this work was completed while Professor Donnelly was visiting the Isaac Newton Institute for Mathematical Sciences at the University of Cambridge. Their support and hospitality are gratefully acknowledged.

** Professor of Law, University of Michigan. B.A. 1973, J.D. 1976, Harvard; D.Phil. 1979, Oxford. — Ed. The authors wish to thank Ian Evett, Peter Gill, Sam Gross, M. Dawn Herkenham, Nancy King, and Paul Robinson for helping in various ways.

1. 509 U.S. 579 (1993).
2. 118 S. Ct. 512 (1997).
3. 293 F. 1013 (D.C. Cir. 1923).

suggested approach sheds a helpful light on a case now pending before the Court, *Carmichael v. Samyang Tire, Inc.*⁴

In Part I of this Article, we summarize the nature of DNA evidence and explain the database search problem. We contrast two types of cases. In both, a sample containing DNA has been left, assertedly by the defendant, at the scene of the crime or some other material location. And in both, the defendant's DNA matches that of the crime sample. But the two cases are different in at least one critical respect. In the first case, what we will call the "confirmation case," other evidence has made the defendant a suspect and so warranted testing his DNA. In the other, what we will call the "trawl case," the DNA match itself made the defendant a suspect, and the match was discovered only by searching through a database of previously obtained DNA samples.

The confirmation case has been the model for most uses of DNA evidence up to now. But the creation of broad databases containing profiles of past offenders, or even of those suspected of crime, means that the trawl case will become increasingly important over the coming years.⁵ The development of such a database is quite advanced in England and Wales. As of October 1998, the English database held 360,000 entries, had matched 28,000 people to crime scenes, and had made 6,000 links between crime scenes.⁶ In the last few years, such databases have developed rapidly in the United States as well, and recently, to considerable fanfare, the FBI began operation of what is in effect a national database.⁷

In Part II, we discuss the NRC's analyses of the question of whether and how the DNA evidence should be presented in the trawl case, and we present our own contrasting analysis. The first NRC report ("NRC I")⁸ opposed admission of the evidence alto-

4. 131 F.3d 1433 (11th Cir. 1997), *cert. granted sub nom. Kumho Tire Co., Ltd. v. Carmichael*, 118 S. Ct. 2339 (1998).

5. See Carey Goldberg, *DNA Databanks Giving Police a Powerful Weapon, and Critics*, N.Y. TIMES, Feb. 19, 1998, at A1:

Of all the new thrills that DNA analysis offers forensic scientists, nothing seems to beat what they call a "cold hit": when a computer discovers the identity of a killer or rapist by matching DNA from blood, semen or saliva left at a crime scene with a DNA profile in a database. A criminal is fingered by his own genes.

Until now, cold hits have come sporadically, mainly in several states where DNA forensic work is most advanced, totaling about 200 nationwide. But Federal and state experts say they will soon be cropping up much more often.

6. See Nicholas Wade, *F.B.I. Set to Open Its DNA Database for Fighting Crime*, N.Y. TIMES, Oct. 12, 1998, at A1; see also *infra* text accompanying note 27.

7. See Wade, *supra* note 6.

8. COMMITTEE ON DNA TECHNOLOGY IN FORENSIC SCIENCE, NATIONAL RESEARCH COUNCIL, DNA TECHNOLOGY IN FORENSIC SCIENCE (1992) [hereinafter NRC I].

gether. NRC I proposed that a match discovered after a database search should provide the basis for performing further tests on a new sample taken from the suspect identified by the match, but that the evidence of the initial match should not be presented to the jury.⁹ The second NRC report ("NRC II")¹⁰ takes a substantially more generous view, but still calls for a testifying expert to drastically understate the value of the evidence.

In our view, the caution reflected in the NRC reports and in the analyses of those who have taken a similar perspective on the problem¹¹ is unwarranted. The fact that the DNA match was found only after a search, possibly of many thousands of samples, does not diminish the value of the evidence; on the contrary, the fact that other potential suspects have been eliminated as possible sources of the crime sample slightly raises the value of the evidence. No downward adjustment in the force of the evidence is appropriate. Indeed, we will show that the confirmation case and the trawl case are really poles of a continuum, and that the analysis of the two cases is fundamentally the same. It may well be that the *total* weight of the evidence is less in the trawl case than in the confirmation case — both because there may be less inculpatory evidence in the trawl case¹² and because the identification of the defendant in that case may taint any later-developed evidence by suggestiveness. But this is an entirely different matter, and one that can be left to the ordinary process of argument to the jury; it does not require any rules restricting the force ascribed to the DNA evidence or the manner in which that evidence is presented.¹³

9. *Id.* at 124, 129.

10. COMMITTEE ON DNA FORENSIC SCIENCE, NATIONAL RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE (1996) [hereinafter NRC II].

11. See N.E. Morton, *The Forensic DNA Endgame*, 37 JURIMETRICS J. 477, 487-92 (1997); Anders Stockmart, *Likelihood ratios for evaluating DNA evidence when the suspect is found through a database search* (Research Rept. 98/1, Dept. of Biostatistics, University of Copenhagen) (to be published, in slightly revised form, in *Biometrics* (Sept. 1999)); Aidan Sudbury, Comment to David J. Balding & Peter Donnelly, *Inference in Forensic Identification*, 158 J. ROYAL STAT. SOC. A, Part 1, at 21, 48-49 (1995); cf. Richard Lempert, *After the DNA Wars: Skirmishing with NRC II*, 37 JURIMETRICS J. 439, 461-62 (1997) [hereinafter Lempert, *After the DNA Wars*] (endorsing NRC I and criticizing NRC II with respect to database searches); Richard Lempert, *Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to the Reverend Bayes*, 13 CARDOZO L. REV. 303, 333-34 (1991) [hereinafter Lempert, *Some Caveats*] (anticipating NRC I and proposing retesting requirement).

12. Indeed, it may be that most of the non-DNA evidence in the trawl case is exculpatory. See, e.g., *Regina v. Adams*, [1998] 1 Crim. App. 377 (Eng. C.A. 1997), discussed *infra* note 74.

13. The arguments made here are consistent with, but fuller in some respects than, those made in IAN W. EVETT & BRUCE S. WEIR, *INTERPRETING DNA EVIDENCE: STATISTICAL GENETICS FOR FORENSIC SCIENTISTS* 219-22 (1998); David J. Balding, *Errors and Misunderstandings in the Second NRC Report*, 37 JURIMETRICS J. 469 (1997); David J. Balding & Peter Donnelly, *Evaluating DNA Profile Evidence When the Suspect Is Identified Through a*

In Part III, we argue that the difficulty manifested by the NRC reports has arisen in part from the tendency of statisticians to export to the legal context methods that were developed to assist scientific inquiries and that appear more suitable in that context than in adjudication. But part of the problem also arises from the tendency of courts to defer to the scientific establishment with respect to matters of scientific evidence. We suggest that the solution lies less in a “gatekeeping” role of the type prescribed by *Frye*, as well as by *Daubert* and *Joiner*, designed to keep out disreputable evidence purporting to be scientific, and more in the role of aggressive consumer. When courts allow experts to present evidence in court, it is to perform a service for the legal system. The courts should try to ensure that the experts are doing so in a way designed to serve the needs of that system and not bound by their own professional habits.

I. DNA EVIDENCE AND THE DATABASE SEARCH PROBLEM

To understand the problem addressed by this Article, it is necessary to understand some aspects of DNA evidence — but, mercifully, only some of the less technical ones.¹⁴

A. *The Nature of DNA Evidence*

DNA is a remarkably complex type of molecule that is sometimes said to contain the genetic blueprint of life. DNA is contained in the nucleus of virtually all cells of every living organism, and within a given organism, the DNA is effectively the same from cell to cell.¹⁵ The entirety of an individual organism’s DNA is referred to as its genome. In humans, the genome consists of two collections, one inherited from each parent, of about three billion building blocks, called *bases*.¹⁶ Human DNA is extremely similar

Database Search, 41 J. FORENSIC SCI. 603 (1996) [hereinafter Balding & Donnelly, *Evaluating DNA Profile Evidence*]; David J. Balding & Peter Donnelly, *Inference in Forensic Identification*, 158 J. ROYAL STAT. SOC. A, Part 1, at 21, 29 (1995) [hereinafter Balding & Donnelly, *Inference in Forensic Identification*]; A.P. Dawid & J. Mortera, *Coherent Analysis of Forensic Identification Evidence*, 58 J. ROYAL STAT. SOC. B, 425 (1996).

14. For helpful descriptions of the underlying genetics of DNA evidence, see NRC II, *supra* note 10, at 12-14, 70-74.

15. Sperm and eggs contain only half the DNA that is in an organism’s other cells.

16. The description here applies only to nuclear DNA. Mitochondrial DNA, found in the energy-producing material surrounding the nucleus, is inherited solely from the mother, and (apart from possible mutations) all offspring of the same mother share the same mitochondrial DNA. For this reason, among others, most forensic DNA testing is done on nuclear DNA, though mitochondrial DNA can be useful when the trace evidence has little or no nuclear DNA. See Mark Hansen, *A Comeback for Hair Evidence*, A.B.A. J., May 1998, at 66.

across individuals. This shared genetic material is what makes us human and distinguishes us from other life forms. At a multitude of sites in the genome, however, there are variations from one human to another; typically, two unrelated individuals will differ at about one site in a thousand. These variations are what make humans genetically different from each other. Except for identical twins, no two humans have DNA that is identical throughout the whole genome.

The consistency of DNA throughout a given person's body, and the uniqueness of a given person's DNA, are what make DNA evidence so valuable for identification purposes. In determining whether the DNA from two separate samples comes from the same person, it is not possible given the current state of science to compare them over the whole genome. Current testing techniques use several *markers*. Each marker targets a particular place, or *locus*, on the genome. For DNA profiling techniques, loci are chosen that display considerable variability among individuals. In most current methods, this variability is manifested by differences in the length, measured by the number of bases or the number of times a given sequence repeats, between pre-specified locations. This procedure will yield two measurements for each sample for each locus, one for the father's side and one for the mother's side.¹⁷

In most modern profiling systems — including the systems used in the databases that are of concern here — the length of DNA fragments can be measured precisely. In comparing two samples at a given locus, if the pair of measurements from one sample is the same as the pair of measurements from the other, the profiles are said to match at that locus; otherwise, they are said not to match at that locus.¹⁸ If the two profiles match at *each* of the loci examined, the profiles are said to match. If the profiles fail to match at one or more loci, then the profiles do not match, and it is virtually certain — putting aside, as we do throughout this Article, the possibility of laboratory error¹⁹—that the samples do not come from the same person.

17. If, however, the individual is *homozygous* at a given locus, meaning that the measured fragments from the mother's and father's sides are the same length, then in some profiling systems only a single measurement, of the common fragment length, will be returned.

18. In older systems, length measurements are only approximate, and the profiles are said to match at a given locus if the pairs of measurements from each are sufficiently close — that is, if they fall within some pre-specified tolerance of each other. There is thus a slight chance that, even if two samples from the same source are measured properly, the test results will nevertheless be sufficiently different that the samples will be declared not to match.

19. This possibility can never be eliminated. We discuss lab error further below. See *infra* note 57.

Our concern here is with the case in which the profiles do match.²⁰ A match does *not* mean that the two samples must absolutely have come from the same source; all that can be said is that, so far as the test was able to determine, the two samples were identical, but it is possible for more than one person to have the same profile as indicated by a test even of several loci. At any given locus, the percentage of people having DNA fragments of a given length, in terms of base pairs, is small but not infinitesimal. DNA tests gain their power from the conjunction of matches at each of several loci; it is extremely rare for two samples taken from unrelated individuals to show such congruence over many loci.

But just how rare? That question must be addressed if the strength of the DNA match is to be assessed. Databases of DNA samples from various populations have been collected, and from these it is possible to estimate how common any given fragment length is at a given locus. It is typically assumed that the measurements yielded by each of the markers used in forensic DNA profiling are independent of one another.²¹ This assumption enables a forensic scientist to multiply probabilities. This multiplication can yield very low estimates of the probability that a given innocent person from the demographic group described by the database would have a DNA profile matching the profile common to both of these samples. Figures in the range of one in millions down to one in many billions are typical of profiling systems now in use.

So far so good; there are complexities and controversies in the process we have described,²² but they are not our concern here, and

20. There are theoretical objections to this aspect of the standard approach in the systems that depend on approximate measurement, for the criteria for determining a match are necessarily arbitrary. See, e.g., BERNARD ROBERTSON & G.A. VIGNAUX, *INTERPRETING EVIDENCE: EVALUATING FORENSIC SCIENCE IN THE COURTROOM* 115-20 (1995); D.A. BERRY et al., *Statistical Inference in Crime Investigations Using Deoxyribonucleic Acid Profiling*, 41 *APPLIED STAT.* 499 (1992). A more theoretically satisfactory approach might be to present the likelihood ratio of the evidence — the ratio of the probability that the actual test results would have occurred if the two samples came from the same source to the probability that the results would have arisen if someone else from a given population were the source of the sample of unknown origin. But such an approach presents practical difficulties, not the least of which is that it may be baffling to a jury. We are inclined to believe that the binary approach, either declaring a match or not according to previously prescribed criteria, is a satisfactory method of summarizing the data, even though some information is lost in this way. In any event, this is the way the data is usually presented, and we will not challenge it here.

21. See, e.g., NRC II, *supra* note 10, at 97-108 (arguing that this assumption of independence may not be precisely accurate, but that if the markers are chosen carefully, it is accurate enough for use in forensic casework).

22. See, e.g., Lempert, *After the DNA Wars*, *supra* note 11; Lempert, *Some Caveats*, *supra* note 11; Kathryn Roeder, *DNA Fingerprinting: A Review of the Controversy*, 9 *STAT. SCI.* 222 (1994). See generally NRC I, *supra* note 8, *passim*; NRC II, *supra* note 10, *passim*.

for the most part this process is by now rather well-accepted. Our concern is with the process by which the matching samples came to be tested, and the implications that this has for the value of the evidence and the manner in which it should be presented.

DNA can be used to test for identity in various contexts in litigation, but we will focus on the most important setting. A crime has been committed, and a person, assertedly the perpetrator, has left a sample of fluid or tissue containing testable DNA at the scene of the crime or at some other scene associated with the crime. This sample is often known as the "crime sample." The police know that if they find a person whose DNA matches that of the crime sample they may have found the perpetrator. Now we will consider two scenarios by which the police might find a match. In each scenario, we will call the person whose profile matches the crime sample "Matcher."

B. *The Confirmation Case*

In the first scenario, what we will call the confirmation case, there is a substantial amount of evidence pointing to Matcher before his DNA is even tested. This evidence might include testimony by a victim of the crime or some other eyewitness identifying Matcher as the perpetrator. It might also include a trail of blood or other circumstantial evidence leading, literally or figuratively, from the crime to Matcher. In any event, the police, believing that they may have their man, secure a DNA test of Matcher. Sure enough, the "suspect sample," as it is often known, matches the crime sample.

Just how this evidence should be presented is a difficult question.²³ One method that is widely used, and that at least for present purposes we regard as satisfactory, states the *match probability*. This is the probability that, if nothing were known about a person other than that he was a member of some defined population, his DNA profile would match that of the crime sample.²⁴ This probability quantifies the rareness of an "innocent" or "chance" match. A forensic scientist testifying in this manner might say something like:

23. See, e.g., *Regina v. Doheny*, [1997] 1 Crim. App. 369, 375 (Eng. C.A. 1996); EVERT & WEIR, *supra* note 13, at 222-38; Richard D. Friedman, *Assessing Evidence*, 94 MICH. L. REV. 1810, 1835-37 (1996); Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence at Trial*, 34 JURIMETRICS J. 21 (1993).

24. As is most common, we use the term match probability to refer to the probability of a match as assessed after the profile of the crime sample is known.

The crime sample matches the sample taken from Matcher. The probability that a sample taken from an arbitrarily chosen member of Defined Population would match the crime sample is 1 in R .

Note that "1 in R " is another way of saying p , where p is the match probability.²⁵

An alternative approach is for the forensic scientist to testify in terms of the *likelihood ratio*. The likelihood ratio of a piece of evidence with respect to two hypotheses is the ratio of the probability that the evidence would arise given the first hypothesis to the probability that the evidence would arise given the second hypothesis. In any given situation, the greater the likelihood ratio, the more strongly the piece of evidence is probative of the first hypothesis as compared to the second. The relationship is described by Bayes's Theorem: The prior odds times the likelihood ratio equal the posterior odds. The prior odds equal the ratio of the probability of hypothesis 1 to the probability of hypothesis 2, assessed in the absence of the piece of evidence in question; the posterior odds are the same ratio, but calculated with that item of evidence taken into account. A forensic scientist testifying in terms of the likelihood ratio might say something like:

These results are R times more likely to have arisen given the proposition that Matcher was the source of the crime sample than given the proposition that an arbitrarily chosen member of Defined Population was the source of that sample.

The likelihood ratio approach may be more satisfying intellectually than the match probability approach, but in the typical confirmation case it is probably less satisfactory practically, in part because it seems to tend to confuse jurors.²⁶

C. *The Trawl Case*

In the second scenario — what we are calling the "trawl case" — apart from the crime sample the police do not at first have evidence that narrows their search to one suspect, or even to a few. This might, for instance, be a rape case in which the rapist was a stranger to the victim and was not apprehended immediately after the crime. But the police do have the crime sample. And they also have a database containing profiles of DNA samples taken from a

25. Because p is a number between zero and one, presumably so small that it has several zeroes immediately to the right of the decimal point, it is often easier to deal with R , a number greater than one.

26. See Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859, 877-80 (1996). See generally *Doheny*, 1 Crim. App. 369 (Eng. C.A. 1996).

large number of people, one of whom *might* be the perpetrator. Sure enough, the database search yields one profile, and only one, that matches the crime sample. The police then try to find further evidence incriminating the source of that profile, Matcher, and perhaps they find it. Perhaps, for example, when Matcher is brought before the victim, she identifies him as the perpetrator.

The trawl case will take on increasing importance in coming years. Development has been fast in England and Wales, where since 1995 DNA profiles have been taken routinely from, among others, all persons charged with a "recordable offence."²⁷ In the United States, development of a national database has been much more complex. In part because of the encouragement, including financial incentives, offered by the DNA Identification Act of 1994,²⁸ all fifty states now require designated sets of convicted offenders to provide DNA samples for analysis.²⁹ The state databases operate in accordance with national quality assurance standards

27. The statutory authority is the Police and Criminal Evidence Act, 1984, ch. 63(3)(b) (Eng.). If the person is subsequently found innocent, in most circumstances the profile is removed from the database. The English database now contains approximately 350,000 samples and is growing at a rate of 200,000 samples per year. Personal communication with Dr. Peter Gill, Forensic Science Service (August 1998).

28. DNA Identification Act of 1994, Pub. L. No. 103-322, §§ 210301-210306, 108 Stat. 2065 (codified at 42 U.S.C. §§ 3796kk-6, 14131-14134 (1994) and codified as amended at 42 U.S.C. §§ 3751, 3753, 3793, 3797 (1994)); see also Harold J. Krent, *Of Diaries and Data Banks: Use Restrictions Under the Fourth Amendment*, 74 Tex. L. Rev. 49, 86 & nn. 173-74 (1995).

29. All but one of the citations are provided in *Statutes and Legislation Regarding Mandatory Submission of Blood Samples for DNA Identification Purposes*, in FORENSIC DNA STATUTES 50 (assembled by Federal Bureau of Investigation, 1998); see also Krent, *supra* note 28.

Most of the state laws require samples only from persons convicted of a narrow set of felonies, including homicide and sexual assault. Four states (Alabama, New Mexico, Virginia, and Wyoming) require samples from all convicted felons. See Wade, *supra* note 6. Louisiana requires samples from, among others, persons *arrested* for felony sex offenses or other specified offenses after September 1, 1999. See LA. REV. STAT. ANN. § 15:609(A) (West Supp. 1999) (effective September 1, 1999); Wade, *supra* note 6.

So far, the constitutionality of the state laws has largely been upheld. See, e.g., Schlicher v. Peters, 103 F.3d 940 (10th Cir. 1996) (upholding a Kansas law requiring persons convicted of designated felonies to submit blood and saliva samples for compilation of a database of genetic markers); Boling v. Romer, 101 F.3d 1336 (10th Cir. 1996) (upholding a Colorado statute requiring an inmate convicted of an offense involving sexual assault to provide the state with a DNA sample before release on parole); Rise v. Oregon, 59 F.3d 1556 (9th Cir. 1995) (upholding a state law that requires persons convicted of murder, a sexual offense, or conspiracy or attempt to commit a sexual offense to submit a blood sample for preparation of a DNA database), *cert. denied*, 517 U.S. 1160 (1996); Jones v. Murray, 962 F.2d 302 (4th Cir. 1992) (upholding a Virginia statute requiring convicted felons to submit blood samples for creation of a DNA database, but holding part of the enforcement mechanism invalid as an *ex post facto* law). *But see* Landry v. Harshbarger, No. CIV.A.98-462 (Mass. Super. Ct., Aug. 12, 1998) (unpublished order preliminarily enjoining enforcement of a Massachusetts statute requiring any person convicted of any of an array of enumerated offenses to submit a DNA sample).

and software designed by the FBI and participate in the Combined DNA Information System (CODIS) maintained by the FBI. Since October 1998, the FBI has been able to compare the profile of a DNA sample from a crime scene with all the profiles in the system. Thus, CODIS now operates in effect as a national database.³⁰ As of October 1998, laboratories reported analyzing more than 250,000 profiles for entry in CODIS, with several hundred thousand more waiting to be analyzed.³¹ As of the same time, CODIS had been responsible for over 400 "hits."³² This number will almost certainly increase dramatically within the next few years.³³

Over time, one can imagine databases even broader than the current ones; fingerprints are now routinely taken from arrestees as well as convicts, and from many persons not suspected of crime, and it is plausible to suppose that the same will occur with respect to DNA samples.³⁴ Indeed, the manager of the English database — which operates under fewer administrative and constitutional constraints than its counterpart in the United States — has said he expects that eventually it will include a third of all English men between the ages of sixteen and thirty.³⁵

And now we can perceive what we have called the database search problem. In the confirmation case, the fact that Matcher's sample, the only one tested, matched the crime sample is clearly powerful evidence that Matcher was the source of the crime sample. But in the trawl case, many samples were tested, without a finger already pointing to any particular suspect. How does this factor affect the strength of the evidence of a DNA match? How, if at all, should that evidence be presented in court? We turn to that problem in Part II.

30. See Wade, *supra* note 6.

31. Personal communication with M. Dawn Herkenham, Chief, Forensic Science Systems Unit, Federal Bureau of Investigation (Aug. 12 & Nov. 3, 1998).

32. *Id.* on Nov. 2, 1998. See also the web page of the FBI's Forensic Science Research and Training Center (visited Dec. 4, 1998) <<http://www.fbi.gov/lab/report/research.htm>>.

33. Note the comment made by NRC II in 1996:

As the number and size of DNA databanks increase, the identification of suspects by this means will become more common. Already, more than 20 suspects have been identified by searches through databases maintained by various states. The number and sizes of these databases are sure to increase.

NRC II, *supra* note 10, at 134.

34. See Wade, *supra* note 6, at A1 (quoting M. Dawn Herkenham, chief of the FBI's Forensic Science Systems Unit: "I think the trend is that 10 years from now all felonies will be covered . . . We recommend that all violent felonies, burglaries, juveniles and retroactivity for people on parole be included.")

Whether such broad databases are appropriate raises issues of policy and constitutional law on which we do not mean to express an opinion.

35. See Wade, *supra* note 6.

II. ANALYZING THE DATABASE SEARCH PROBLEM

In Section A of this Part, we summarize the analyses of the database search problem in NRC I and NRC II. Section B offers our basic analysis. Section C then re-examines the NRC analyses, focusing on NRC II and demonstrating what we believe to be its flaws. Section D discusses how evidence of a DNA match after a database trawl fits into the case as a whole. Section E addresses the issue of how such evidence should be presented to the jury. And Section F criticizes the recommendation for re-testing made by NRC I and recently endorsed by some scholars.

A. *The NRC Analyses*

1. *NRC I*

In NRC I, a Committee on DNA Technology in Forensic Science specially appointed by the NRC said:

The distinction between finding a match between an evidence sample and a suspect sample and finding a match between an evidence sample and one of many entries in a DNA profile database is important. The chance of finding a match in the second case is considerably higher, because one does not start with a single hypothesis to test (i.e., that the evidence was left by a particular suspect), but instead fishes through the databank, trying out many hypotheses.³⁶

Thus, the Committee pointed out, “[i]f a pattern has a frequency of 1 in 10,000, there would still be a considerable probability (about 10%) of seeing it by chance in a databank of 1,000 people.”³⁷ The report recommended a cautious solution: A match between an “evidence sample” — what we are calling the crime sample — and a profile in a database “should be used only as the basis for further testing using markers at additional loci.”³⁸ That is, evidence of the initial match should be deemed “probable cause” for securing a blood sample from the person so identified, and comparing it with the evidence sample using markers that were not used in the initial test. What is more, if the second test indicates that the two samples matched, “only the statistical frequency associated with the additional loci should be presented at trial (to prevent the selection bias that is inherent in searching a databank).”³⁹

36. NRC I, *supra* note 8, at 124.

37. *Id.*

38. *Id.* at 129.

39. *Id.* at 124.

To understand this approach, which has recently been endorsed by eminent scholars,⁴⁰ consider a simple example. Suppose that a trawl through a database containing 100,000 (or 10^5) profiles revealed that Matcher's profile, and no others in the database, matched the crime sample, and that the probability that the profile of an arbitrarily chosen member of Defined Population would match the crime sample is 1 in 10,000,000 (or 1 in 10^7). This in itself would not be admissible evidence under the approach recommended by NRC I. But it would justify taking another sample from Matcher and using additional markers to compare it to the crime sample. Suppose further that on this second test Matcher's profile matches that of the crime sample, that the match probability on this second test is 1 in 1,000,000 (or 1 in 10^6), and that the markers used in this second test are independent not only of each other but of those used in the first test. Then, according to NRC I, the prosecution should be allowed to present evidence that Matcher's DNA matches that of the crime sample. But the prosecution's forensic science witness should not be able to testify that the probability of a match of the crime sample and the DNA of an arbitrarily chosen member of the population is 1 in 10 trillion (or 1 in 10^{13}) — which is the probability, assuming the results of the two tests to be independent, that an arbitrarily chosen member of Defined Population would match the crime sample on all the markers used in both tests. Instead, the witness would only be allowed to use the figures for the second test, reporting that the probability of a match is 1 in 10^6 .

2. NRC II

NRC II offered an analysis of the database search problem that was quite similar to that of NRC I, but recommended a substantially more lenient solution. According to NRC II, an "important difference," a difference of a "logical" nature, between the confirmation case and the trawl case is illustrated by this simple set of statistical facts:

[I]f we toss 20 reputedly unbiased coins once each, there is roughly one chance in a million that all 20 will show heads. According to standard statistical logic, the occurrence of this highly unlikely event would be regarded as evidence discrediting the hypothesis that the coins are unbiased. But if we repeat this experiment of 20 tosses a large enough number of times, there will be a high probability that all

40. See *infra* notes 81-82 and accompanying text (views of Professors Morton and Lempert).

20 coins will show heads in at least one experiment. In that case, an event of 20 heads would not be unusual and would not in itself be judged as evidence that the coins are biased.⁴¹

Further, contended the report, “[t]he initial identification of a suspect through a search of a DNA database is analogous to performing the coin-toss experiment many times: A match by chance alone is more likely the larger the number of profiles examined.”⁴²

Thus, NRC II indicated that it is necessary “[t]o decide if the database search itself has contributed to obtaining a match,” in the same way that the repetition of the twenty-coin experiment might have been “responsible for producing the 20 heads.”⁴³ The report therefore argued that the appropriate quantity to consider is the probability, under the hypothesis that the crime sample “was not left by someone whose DNA profile is in the database (or a close relative of such a person),” that at least one of the database profiles would match the crime sample.⁴⁴ Assume, as is usually the case, that p , the probability that any given profile will match that of the crime sample, is very small. Then, making some other simplifying assumptions, the probability that at least one profile from a database of $D + 1$ profiles will match the crime sample is very close to, and no greater than, $(D + 1)p$.⁴⁵ That is, each profile compared increases the probability that there will be a match by a tiny amount, very close to p . $D + 1$ comparisons therefore yield a probability of just about — but not exactly! — $(D + 1)p$.⁴⁶

Accordingly, NRC II recommended that “[w]hen the suspect is found by a search of DNA databases, the random-match probability should be multiplied by . . . the number of persons in the database.”⁴⁷ Thus, instead of using R , which equals $1/p$, the forensic scientist would use

$$R/(D + 1). \quad (1)$$

If testifying in terms of the match probability, she would say,

The probability that a sample taken from an arbitrarily chosen member of Defined Population would match the crime sample is 1 in $R/(D + 1)$.

41. NRC II, *supra* note 10, at 134.

42. *Id.*

43. *Id.*

44. *See id.*

45. *Id.* at 135.

46. The exact formula for the probability of at least one match is $1 - (1 - p)^{(D+1)}$. If $D + 1$ is much smaller than $1/p$, which we have called R , then this quantity is very close to $(D + 1)p$.

47. Recommendation 5.1, NRC II, *supra* note 10, at 40.

We will refer to this argument, to which NRC II gives primacy, as its *frequentist argument*.

NRC II also presents another, distinct, argument, which we will call its *likelihood ratio argument*. "If one wishes to describe the impact of the DNA evidence under the hypothesis that the source of the evidence sample is someone in the database," recommends NRC II, "then the likelihood ratio should be divided by [the number of profiles in the database.]"⁴⁸ That is, according to this argument, the likelihood ratio of evidence of a single match found after searching a database of $D + 1$ profiles is not R but rather $R/(D + 1)$,⁴⁹ which we have already seen as Expression (1).

The two arguments in NRC II are different. The frequentist argument assesses the probability that *at least* one profile in the database would match the crime sample by chance, while the likelihood ratio argument attempts, by different means, to assess the value of the evidence that *exactly* one of those profiles matches the crime scene. These approaches lead to different answers.⁵⁰ The factor of $(D + 1)$ — by which, according to the report, the probative force of the evidence should be reduced to account for the fact that the match resulted from a database trawl — is a conservative approximation under the frequentist approach, but an exact number under the likelihood ratio approach. Both approaches, though, share the perception that the more profiles examined, the less probative the evidence.

B. *Our Analysis*

Though the NRC reports differ in their ultimate recommendations, their analyses of the database search problem are very similar. We believe that these analyses, and those of scholars who have supported the NRC approach, are clearly wrong. They ask the wrong question, and they fail to recognize the full import of evidence of identification based on a database search.

The proper view of the situation, which we will present here, reflects a rather simple intuition. The value of a DNA match is attributable to the rarity of the profile. If the DNA of a particular person matches the crime sample, that evidence strongly supports

48. NRC II, *supra* note 10, at 40. The argument is developed in detail in Stockmarr, *supra* note 11, at 6-9.

49. We show how to derive this ratio in the Appendix. See also Stockmarr, *supra* note 11, at 8.

50. These differences are small when the databases are small, and they disappear altogether in the limiting case in which there is only one profile in the database.

the proposition that *that person* was the source of the crime sample; that is, the evidence makes that proposition appear far more probable than it did before the match was known. That other samples have been tested and found not to match does not weaken the probative value of the match, with respect to *this particular proposition*, which is the one of interest at the time of trial. On the contrary, this result somewhat strengthens the probative value of the match, because it eliminates some other persons as potential sources. How probable it appears that the particular person is the source depends not only on the DNA evidence but also on the other evidence in the case. If there is no other evidence pointing to him, then the proposition will not appear as likely as if there were such evidence — not because the DNA evidence is any less valuable, but because the prior probability of the proposition is so low. And evidence found after the DNA match is determined might be subject to a ground of skepticism — the possibility of suggestiveness created by the match itself — not applicable to evidence found beforehand. Thus, the probability that the defendant is the source of the crime sample may well appear less in the trawl case than in the confirmation case, but this is not because the DNA evidence itself is any weaker in the trawl case.

We will now explore the reasoning that leads to these conclusions.

Both NRC I and NRC II emphasized that, as the number of profiles tested increases, so too does the probability of finding a match with the crime sample. That is indisputably true. One can even say that the larger a database is the more likely it is that the database will yield at least one false positive result — a profile that matches the crime scene sample but that does not come from the source of that sample.⁵¹ But the conclusion that the NRC reports draw is that the larger a database is (up to a point) the less valuable

51. This appears rather obvious: The more profiles there are in the database, the more probable it is that at least one of them will match that of the crime sample but be from someone other than the source of that sample. That is, the number of false positives will tend to increase with the size of the database.

It is even true that the larger the database is, up to a point, the more probable it is that there will be *exactly* one false positive. (Our analysis has focused, for simplicity, on the most important case, in which the database search turns up exactly one match with the crime sample.) One attempting to support the NRC analyses might well point out that, the larger the database is, up to a point, the more probable it is that there will be exactly one false positive. That is, the more probable it is that both the following propositions are true: (1) A search of the database will turn up exactly one sample that matches the crime sample, *and* (2) in fact the source of the crime sample is someone who has not contributed a sample to the database. After a point, this conjunction of possibilities becomes less likely because the second part of it — that the source was someone outside the database — becomes very unlikely, and the first part may also become less likely.

is evidence that a database trawl yielded a single match. Here the NRC and its supporters go wrong.

The proposition that the DNA evidence is offered to prove is not the broad one that the source of the crime sample is a person represented in the database. Rather, it is that *one particular person — the defendant in the case at hand — is the source of that sample*. And the evidence bearing on this proposition is not simply that there was one match within the database. Rather, it is that *the DNA of that particular person — alone of all those tested — matches the crime sample*. We will examine in turn each of these assertions.

1. *Identifying the Material Proposition*

First, consider the material proposition — or, in the words of Federal Rule of Evidence 401, the “fact that is of consequence to the determination of the action” — that the DNA evidence is offered to prove.⁵² It is the guilt of the defendant, whom we have called Matcher, that the prosecution wants to prove; that is what the case is about, and all the prosecution’s evidence is offered to raise the probability of the proposition that *the defendant* is guilty. The DNA evidence cannot answer whether the defendant committed the crime alleged. But if Matcher was the source of the crime sample, then Matcher’s guilt presumably appears much more likely. It is the probability that Matcher is the source of the crime sample that is of importance to the prosecution — not the larger probability that *somebody* represented in the database is the source.

Adherents to the NRC approach believe there is a flaw in this reasoning. The material proposition on which DNA evidence bears, in our view, is that Matcher is the source of the crime sample. How can we identify that as the material proposition if there was no evidence pointing to Matcher — he was not a suspect — before the database search singled him out for attention?⁵³ We will return to this question in Part III, where we will analyze the statistical way of thinking that appears to make it troublesome. For now, we will offer an answer that satisfies us.

Even if at any stage of the investigation or trial there is not yet any evidence pointing to Matcher, he, like everybody else in the world who had not been eliminated as a possibility, is in fact a sus-

52. See FED. R. EVID. 401 (defining “relevant evidence” as “evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence”).

53. See, e.g., Stockmarr, *supra* note 11, at 6-8.

pect in the limited sense that he is *possibly* the source of the crime sample. Thus, an investigator or factfinder asked to do so might assign a probability to the proposition that Matcher is the source. The absence of other evidence pointing to him would presumably yield prior odds in favor of the proposition — that is, the odds as assessed without benefit of the evidence of interest, here the DNA evidence — that were extremely low. But those odds would be above zero, and so subject to being raised by evidence pointing in Matcher's direction. All other things being equal, the lower the prior odds favoring the proposition that Matcher is the source, the lower the posterior odds — the odds assessed with the DNA evidence taken into account — in favor of the proposition; recall that, according to Bayes's Theorem, the posterior odds of a proposition equal the prior odds of the proposition times the likelihood ratio. In other words, the fact that Matcher is not a particular object of suspicion before the DNA evidence is fully taken into account by the extremely low prior odds.

2. *The Impact of the Evidence*

Having identified the proposition of interest — that Matcher is the source of the crime sample — let us now examine the impact of the evidence at issue. This is proof that *Matcher's* DNA profile matches that of the crime sample *and* that the profiles of none of the other people represented in a database match that sample.

The most important aspect of this evidence is that Matcher's DNA profile matches that of the crime sample. The significance of this fact, with respect to the question of whether Matcher was the source of the crime sample, is unaltered by the fact that other samples have been tested. Assuming hypothetically that Matcher is the source, the probability of *this* match is, for all practical purposes, 1. Assuming hypothetically that Matcher is not the source, the probability of *this* match may be assessed as p , where p is the probability that the DNA of a person arbitrarily selected from the suspect population would match the crime sample;⁵⁴ note that $1/p$

54. The suspect population is the population from which the crime sample is believed to originate. The notion of a suspect population is a useful simplification for some analyses, but its limitations are important to recognize. It seems to suggest that all people within a defined group are equally likely, before any other evidence is received, to have committed the crime, and that anyone outside that group could not have committed it. But clearly this is not so. In the typical stranger rape case, for example, a man who lives within five miles of the crime scene is more likely, absent other evidence, to have committed the crime than a man who lives 300 miles away, and that man is a more likely suspect than one who lives on the other side of the world — but even the man from the antipodes cannot be altogether excluded. For further comments, see EVETT & WEIR, *supra* note 13, at 27-28, 233-34.

equals R , the number used in both the match probability and the likelihood ratio forms of testimony.⁵⁵ Neither of these probabilities is altered by the fact that other samples have been tested, or by the degree to which an investigator believed before testing began that Matcher was a likely suspect.

Now consider in addition the fact that other samples have in fact been tested and found not to match the crime sample. With respect to the precise proposition at issue — that Matcher is the source of the crime sample — this fact can only enhance the probative value of the DNA evidence. One reason for this is that the additional information that a significant number of persons have been tested and found not to match the crime sample can only make the profile of that sample appear rarer than it did absent that information. This factor will almost always be of negligible importance, and we will put it aside for further analysis.⁵⁶ Potentially more importantly, a number of people other than the defendant who previously appeared to be possible sources of the crime sample have now been eliminated, thus making each of the remaining possibilities somewhat more probable. Assuming, as is usually the case, that the size of the database is very small in comparison to the suspect population, this effect too will be negligible, but as the size of the database increases in comparison to that population, the effect becomes dominant. If the database includes the entire suspect population, then the existence of only one match points the finger without doubt (assuming accurate testing)⁵⁷ at the person so identified.

55. See *supra* section I.B.

56. This effect is demonstrated in Balding & Donnelly, *Evaluating DNA Profile Evidence*, *supra* note 13, at 604; EVETT & WEIR, *supra* note 13, at 220.

57. Some perceptive scholars have emphasized the importance of the possibility of laboratory error, which is often far more plausible than the possibility of a match by chance. See, e.g., Jonathan J. Koehler, *Why DNA Likelihood Ratios Should Account for Error (Even When a National Research Council Report Says They Should Not)*, 37 JURIMETRICS J. 425 (1997); Lempert, *After the DNA Wars*, *supra* note 11, at 446-54; William C. Thompson & S. Ford, *The Meaning of a Match: Sources of Ambiguity in the Interpretation of DNA Prints*, in FORENSIC DNA TECHNOLOGY 93 (Mark A. Farley & James J. Harrington eds., 1991). We agree that jurors should take the possibility of lab error into account in their subjective assessment of the evidence. It is important to emphasize, however, that what matters is not the probability of any laboratory error, but rather only the probability of those errors that would lead to the false declaration of a match in the given case — a probability that will vary widely with the circumstances of the DNA testing. For a fuller treatment, see David J. Balding & Peter Donnelly, *Inferring Identity from DNA Profile Evidence*, PROC. NATL. ACAD. SCI. (USA) 11,741 (1995). See also *infra* notes 59-60 and accompanying text (analyzing “paradox of the lottery” in similar terms); A. Philip Dawid & Julia Mortera, *Forensic Identification with Imperfect Evidence*, 85 BIOMETRIKA 835, 844 (1998) (drawing analogy to lottery paradox but inaccurately ascribing erroneous argument to Balding and Donnelly). In particular, the probability of this type of error will tend to be smaller in a trawl case than in many confirmation cases because in the trawl case the DNA profiling of the crime sample and the defendant’s sample are performed at different times and often in different laboratories.

As shown in the Appendix, the basic nature of this relation may be expressed rather simply in symbols. Subject to some simplifying assumptions, the likelihood ratio of the DNA evidence equals

$$R \times [1 + mD/N], \quad (2)$$

where D is the number of people in the database, apart from the one member — here *Matcher* — whose DNA matches that of the crime sample; N represents the number of people in the suspect population but not in the database; and m is an optional multiplier meant to reflect how many times more likely a given person represented in the database (other than *Matcher*) was believed, before the search results were known, to be the source of the crime sample than was a given person in the suspect population but not in the database. This multiplier is optional in the sense that, if it is considered inappropriate to assume that persons represented in the database are more likely than others to be the source of the crime sample, the multiplier can simply be assumed to be 1. Note several points about Expression (2).

(a) Because neither m , D , nor N can be negative, this expression cannot be less than R , the likelihood ratio of a match in the confirmation case. Because p , the match probability, is less (presumably much less) than 1, the likelihood ratio R , which equals $(1/p)$, is greater than 1, meaning that the evidence tends to support the proposition at issue. And if m , D , and N are all positive, then Expression (2) indicates that the likelihood ratio of the DNA evidence yielded by the database search is at least slightly *greater* than that for comparable evidence yielded by a match in the confirmation case.

(b) If D is zero — that is, no DNA is tested but that of the person who yields a match — Expression (2) yields a likelihood ratio of R , just what we would expect for the confirmation case.

Indeed, though for convenience we have been treating the confirmation case and the trawl case as antithetical, we should recognize that they are really variations of a more generalized database search case: In the confirmation case, the database consists of only one profile, instead of many, and typically the investigator has a relatively high prior probability that the person who is the source of that profile is also the source of the crime sample. Moreover, many

In addition to providing information that will assist the jury in assessing the probability of a false report of a match in the case at hand, an expert should provide information bearing on the probability of an actual match on the assumption that the defendant is innocent. We focus in this article on how the latter probability — which is also the focus of most expert testimony in this area — is affected by the fact that the match is the result of a database search.

cases fall in the middle. The police may have several people under suspicion and test them all. In such a case, the profiles of all of the suspects may be considered a relatively small database, and the prior probability for each suspect is smaller than in the confirmation case but much greater than in the trawl case. The facts of each variation are different — but the fundamental analysis is the same.

(c) If N is zero — that is, there is nobody in the suspect population whose DNA is not tested — then Expression (2) yields an infinitely large likelihood ratio. This again is what we would expect. It means that the fact of Matcher's match and the absence of any other match are together effectively absolute proof — putting aside such considerations as laboratory error — that Matcher was the source of the crime sample.

C. *NRC II Revisited*

In Section B, we have introduced the basics of our analysis. With this background, we can examine more closely the logic, or as we see it the illogic, of the NRC analysis. We will focus here on NRC II, which not only is more recent but offers fuller arguments, with two variations, the frequentist and likelihood ratio arguments. We will discuss these arguments both separately — for they are different, the differences having caused some dissension in the NRC camp⁵⁸ — and together, for they share basic flaws.

1. *The Frequentist Argument*

First consider the frequentist argument. Our analysis shows the immateriality of NRC II's hypothetical of an experiment in which twenty coins are repeatedly tossed once each. As NRC II points out, if this experiment is done enough times, an event of twenty heads is not so unusual and does not strongly suggest that the coins are biased. But this hypothetical does not represent the reality of the situation involving DNA evidence. A more material hypothetical would be as follows: A coin known to be biased, but otherwise indistinguishable from all those in a piggy bank, is thrown into the bank. From the coins in the bank, a tester picks out a handful at random, and flips each of them twenty times. Each lands heads up approximately ten times — except for one coin, which shows heads on all twenty flips. This experiment is powerful, albeit not conclusive, evidence that the coin that landed heads up is the biased coin

58. Professor Morton's approach is founded on the frequentist argument. See Morton, *supra* note 11, at 488. Dr. Stockmarr explicitly rejects that argument, giving the likelihood ratio argument instead. See Stockmarr, *supra* note 11, at 9.

that was thrown into the bank. Just how powerful the evidence is will depend on how common biased coins are believed to be in the bank. But one point seems utterly clear: The evidence that the one twenty-heads-up coin is the biased one thrown into the bank is made marginally stronger, not less strong, by the fact that other coins were tried and appeared to be unbiased.

Another analogy might also highlight the failure of the NRC analysis. For many years, skeptics of conventional probability theory have posed the so-called "paradox of the lottery." There is some chance that the person announcing the winning ticket in a lottery will get the number wrong. If the number of tickets in the lottery is huge, the prior probability — assessed before the announcement — of any given ticket being the winner is very small. So if the number of tickets is large enough, does this mean that we should not have much confidence in the announcement of the winning ticket?⁵⁹ No, it does not. The proposition at issue is that a given ticket *T* is the winner, and the evidence presented is that *T* was announced as the winner. If *T* is not in fact the winner, it is not enough to account for that evidence merely that the announcer made a mistake. Rather, the announcer would have had to make *that particular mistake* — announcing *T* among all other possibilities — to produce this evidence. That is extremely unlikely, and the more lottery tickets there are the more unlikely it is. Hence, the announcer's declaration that *T* is the winner carries great weight, as one would expect.⁶⁰ Similarly, the NRC analysis of the database search problem fails to take into account that the proposition at issue at trial is that a particular person was the source of the crime sample, and that the evidence presented is that *that particular person's* DNA matches the crime sample.

2. *The Likelihood Ratio Argument*

Now consider the NRC II likelihood ratio argument. Recall that this argument considers the likelihood ratio of the evidence

59. Two early presentations of the problem are RICHARD PRICE, *FOUR DISSERTATIONS* 410-12 (1767), and ISAAC TODHUNTER, *A HISTORY OF THE MATHEMATICAL THEORY OF PROBABILITY FROM THE TIME OF PASCAL TO THAT OF LAPLACE* 400 (1865). Todhunter's presentation is quoted in Lea Brilmayer & Lewis Kornhauser, *Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116, 148 n.114 (1978). See also L. Jonathan Cohen, *Can Human Irrationality Be Experimentally Demonstrated?* 4 BEHAV. & BRAIN SCI. 317, 329 (1981).

60. For fuller analyses see Richard D. Friedman, *Route Analysis of Credibility and Hearsay*, 96 YALE L.J. 667, 683-85, 736-39 (1987); Jonathan J. Koehler & Daniel Shaviro, *Veridical Verdicts: Increasing Verdict Accuracy Through the Use of Overtly Probabilistic Evidence and Methods*, 75 CORNELL L. REV. 247, 270-71 (1990).

that there is exactly one match in the database, for the hypotheses that the source of the crime sample is, or is not, in the database. This likelihood ratio is $R/(D + 1)$. We have no quarrel with the calculation of this ratio.⁶¹ In our view, NRC II has simply given the right answer to the wrong question. Indeed, our analysis has suggested that the question NRC II answers is wrong in two ways. Its likelihood ratio does not pertain to the evidence at hand (that Matcher's profile matches the crime sample), nor does it concern the proposition at issue (that Matcher is the source of the crime sample). Rather, it has treated the evidence as revealing only the limited fact that within the database there was exactly one match of the crime sample, and has determined the likelihood ratio of this evidence for the generalized proposition that the source of the crime sample is someone reflected in the database.⁶² With respect to that generalized proposition, the evidence does indeed have lower probative value the more profiles there are in the database. But that proposition, as we have seen, is not what is at issue.

3. *Anomalies*

We can now see that the analyses of NRC II lead to various anomalies.

a. The large database. Suppose that the database is large in comparison to the suspect population. This could happen because the suspect population is very small — if, for example, it appears that the crime must have been committed within a closed group such as the people of an isolated village or the crew of a ship at sea. Or, in theory, it could happen if an enormously large database were compiled — if, say, it became accepted practice throughout the world to take a DNA sample from every person at birth. It is obvious that if everyone in the suspect population is represented in the database, then proof that exactly one person's DNA matches the crime sample is conclusive proof (again, putting aside the possibility of lab error) that this person is the source. And this is in accord with the analysis we have presented, which indicates a steady, continuous relation: As indicated by Expression (2), the greater the percentage of the suspect population that is represented in the database, the more probative is evidence that exactly one profile in the database matches the crime sample.

61. We present a derivation of this likelihood ratio in Section D of the Appendix.

62. That proposition would also have higher prior odds the more profiles there are in the database, and these two factors would tend to balance out. *See infra* section II.D.2.

NRC II agrees — as it must, for the point is obvious — that for a sufficiently large database this evidence provides overwhelming proof that the one person identified by the match is the source of the crime sample.⁶³ But how can this concession be reconciled with the report's assertion that the strength of the evidence is weakened by a factor equal to the number of persons in the database? The report states that “[i]f the database searched includes a large proportion of the population, the analysis must take that into account.”⁶⁴ It gives no explicit advice as to how this can be done, however — for the simple reason that it cannot be done without undermining the underlying argument.⁶⁵

b. The exculpatory match. If the NRC likelihood approach were correct, in some circumstances the following bizarre outcome could arise: A database would be searched, it would turn up one match with the crime sample, and the person so identified could claim that this evidence tends to exculpate him!

63. See NRC II, *supra* note 10, at 40 (“In the extreme case, a search of the whole population should, of course, provide a definitive answer.”)

64. *Id.*

65. Expression (1), which is the key quantity for both of the NRC's arguments, diminishes constantly as the size of the database increases, indicating diminishing value of the evidence. There is no point at which it begins to move in the other direction — obviously so, for the size of the database is the denominator of the Expression and is unrelated to the numerator. NRC II offers no hints at what point, or under what circumstances, the arguments leading to Expression (1) lose force and some other argument becomes dominant.

NRC II does suggest that the likelihood ratio argument leads to a proper assessment of posterior odds, see NRC II, *supra* note 10, at 164-65, but as we indicate *infra* section II.D.2, this is no consolation for the argument.

Stockmarr offers an extended discussion of the problem of the large database, see Stockmarr, *supra* note 11, at 10-14, but his argument is irretrievably flawed. Stockmarr's argument, like the NRC's, appears to be based on the fact that he is able to produce a calculation yielding appropriate posterior odds for the case in which the database trawl yields one match with the crime sample; the argument therefore suffers the same flaws as does the NRC's. Stockmarr calculates the posterior odds for the trawl case (which we show in the Appendix to equal R/N , in our notation), and then divides them by the likelihood ratio for the confirmation case (R in our notation) to yield a factor, equal to the reciprocal of the population not represented in the database (that is, $1/N$ in our notation), that he calls a “sliding correction” to take the size of the database into account. But the “correction” merely means that he is able to reach the correct posterior odds. It is unclear what significance Stockmarr attaches to the factor, given that he still maintains that the same likelihood ratio prevails for the evidence no matter how large the database. Stockmarr's interpretation also produces bizarre results — not surprising, given that he is comparing the posterior odds from the trawl case with the likelihood ratio from the confirmation case, which, to borrow a phrase from the English legal scholar Peter Mirfield, seems to us to be like comparing apples and Tuesdays. Thus, he contends that if the database contains profiles from all but one member of the population the evidence of a match has just the same strength as it would in the confirmation case, and that if the database is at all smaller the evidence is weaker than in that case. See *id.* at 14. He seems not to recognize fully that a single match after the search of such a nearly all-inclusive database would be virtually conclusive evidence against the source of the single matching sample.

How could that occur? Recall that under the NRC likelihood approach, the likelihood ratio of the DNA evidence, given one match with the crime sample, is $R/(D + I)$. But there is no reason why this ratio must be greater than one. In fact, if $D + I$ is greater than R then the ratio is less than one. That could happen if the database (represented by $D + I$) is relatively large or if the DNA profile is relatively common (reflected by a low value for R). Such a profile could occur if the DNA is so degraded that only a few markers yield test results and only a partial profile is obtainable.⁶⁶

The root of the problem is easy to find. It is that the NRC analysis focuses on the wrong question, whether the source of the crime sample is represented in the database, not on the question of whether the defendant is the source of the sample. R is the number of profiles one would expect to have to search to get one match of a given profile. Thus, if the database is larger than R , and yet only one match turns up, that does indeed make it less probable that the source is profiled in the database, because if the source were profiled in the database we would expect to see at least two matches of the crime sample in the database, one from the source and at least one by chance.

The problem is eliminated, then, by focusing on the correct proposition. Whatever the relation between R and $D + I$, the evidence makes it more probable — as our analysis indicates — that *the person identified by the trawl* is the source.

c. *The lucky strike.* Now we will explore a pair of anomalies that expose some ambiguity in the rationale of the NRC analysis. Under that analysis, the value of DNA evidence is degraded when (1) many DNA profiles are compared to that of the crime sample, and (2) no individual was previously a strong suspect. But each of these factors can exist without the other.

Suppose first that the police, with no particular reason to suspect anybody, decide to compare the DNA profile of one person chosen at random from the database — again, call him *Matcher* — with that of the crime sample. And suppose that they get lucky: The profiles happen to match. Plainly, this evidence is no stronger, with respect to the proposition that *that person* was the source of

66. R would also be relatively low if the evidence were of some other kind of profile, such as standard blood groupings, which are not as rare as DNA profiles; the validity of the statistical analysis does not depend on the identification technique used, and nothing in the derivation of the NRC's likelihood ratio (Expression (1)) depends on either the size of the database or the value of the match probability.

Just as it is untenable to adjust the likelihood ratio prescribed by NRC II for large databases, it is untenable to make another adjustment for relatively indiscriminating tests.

the sample, than that which would be yielded if the entire database were searched, and only that same person's profile matched that of the crime sample. Indeed, this evidence is weaker, because unlike the database search the one-to-one comparison does not eliminate other suspects. And so our analysis indicates.

But what would the NRC approach say? If the key question under that approach is whether multiple profiles have been compared with the crime sample profile, then it would treat evidence of this one-on-one comparison as stronger than the evidence in the trawl case. Indeed, it would appear that the NRC should treat this evidence as no weaker than that indicated by a match in the confirmation case. This appears anomalous.

Perhaps, then, the NRC's supporters would say that the DNA evidence is in fact weaker than its one-on-one nature suggests, and that the key factor is that before this evidence there was no particular reason to suspect Matcher. But, as we argue in the next subsection, the suggestion that the correct interpretation of the DNA test depends on the degree of suspicion of Matcher prior to that test is not only logically untenable but also unworkable. Moreover, neither NRC approach provides any method of downgrading the evidence in any case where only one sample was compared with the crime sample.

d. The precautionary search. The "lucky strike" case featured neither a prime suspect before DNA testing nor a database search. Now turn the case around so that it includes both. Suppose — more realistically — that the police suspect that Matcher is the source of the crime sample, but that they compare the profile of that sample not only with Matcher's but also with a database of *D* other profiles, and that only Matcher's matches the profile of the crime sample.

It seems blatantly obvious that this evidence is stronger, not weaker, than the evidence typically yielded in the confirmation case, in which the authorities would have proof of the match between Matcher and the crime sample and nothing more: The search through the database gives the same information as in the single-comparison case, and then some, by eliminating a group of suspects. Once again, the approach we have presented is in accord with the intuition that the more profiles that are compared and fail to match the crime sample, the more probative is the evidence of the one sample that does match.

What would the NRC treatment be? On the face of the NRC analyses, it appears that they would point in the wrong direction,

treating the evidence as weaker the more profiles that are searched and found not to match. If this is true, then a suspect whose DNA is about to be tested has an unambiguous incentive to demand that the crime sample be measured against all available databases.⁶⁷ Assuming the suspect's profile matches that of the crime sample, the probative value of that evidence would obviously be diminished if the database search turned up another match with that sample. But under the NRC logic, that value would also be diminished even if the database search turned out negative!⁶⁸

Once again, a defender of the NRC approach might contend that the key is not how many samples were tested but whether Matcher was a suspect before the DNA tests were done. Here he was such a suspect, the argument would go, and therefore the formulas indicating degradation of probative value when a match is found with a person who was not previously a suspect do not apply.

This argument fares no better. For one thing, it appears to give up most of the case. As we have indicated above, even in the trawl case Matcher was a suspect before the DNA tests were conducted, in the sense that an investigator presumably would have assigned a positive probability to the proposition that Matcher was the source. If the response is that the treatment depends on the *degree* of prior suspicion of Matcher — above a given threshold, the evidence to be treated as if this were the confirmation case, and below as if it were the trawl case — that also creates serious difficulties.⁶⁹ There seems to be no reasoned basis on which to choose such a threshold, nor any reason why slight differences across such a threshold should have such enormous consequences⁷⁰ in the treatment of the evidence. And attempting to determine whether the threshold had been crossed would presumably require a jury assessment of the subjective state of the (possibly collective) mind of the investigator at the time the DNA tests were ordered.

e. The sequential search. Now suppose that the police follow this strategy: They investigate the crime, without using DNA evidence, until they identify a specific suspect as the most likely perpe-

67. See Balding & Donnelly, *Evaluating DNA Profile Evidence*, *supra* note 13, at 605.

68. Indeed, if the database is large enough, the defendant might — under this theory — even be able to make the match appear exculpatory. See *supra* section II.C.3.b.

69. Note that the prior probability that Matcher is the source is less in some confirmation cases than in some trawl cases. The degree of suspicion of Matcher might be very slight in a confirmation case, and in a trawl case it may be that everyone outside a very small database has been eliminated as a suspect.

70. That is, a factor equal to the size of the database in assessing the match probability or the likelihood ratio.

trator. They then compare his DNA profile to that of the crime sample. If the profiles match, he is charged, and no other suspects are considered. If the profiles do not match, the police repeat the process with the person who now appears to be the most likely suspect. The process continues until a suspect matches the crime sample or, for some reason such as limited resources, the investigation is halted.

Though this model may capture the important features of many police investigations,⁷¹ there is no guidance in either NRC report on how to handle it. The logic of the NRC frequentist argument suggests that the quantity of interest to the court is the probability that this search process would result in a match even though all the suspects tested were innocent.⁷² Assessing this probability — which depends on an assessment of the chance, across a wide array of counterfactual circumstances, that the investigation would have been called off before a match was achieved — is extremely difficult and highly subjective. Perhaps more seriously, this analysis leads to an anomaly. Suppose that the crime is sufficiently serious that the police would in all probability continue the search until a match was found. Then under the NRC frequentist approach (and, though it is hard for us to be sure of the NRC position, probably under its likelihood ratio approach as well), the evidence of the match would be nearly worthless.

This cannot be. One person tested, Matcher, has matched the crime scene sample. This evidence is not by any means conclusive, for there may well be other people with the same profile. But, given that the profile is relatively rare it is nonetheless powerful evidence against him, making the proposition that he is the source far more probable than it was before. And this power can only be increased by the fact that others, who apart from the DNA evidence appeared to be more likely suspects, have been eliminated.⁷³

f. Summary. These anomalies all highlight the fact that, because they fail to ask the right question, NRC II and its supporters fail to recognize the full force of proof that a single profile in a

71. Particularly if it is extended to include the possibility that the police simultaneously send samples from several suspects for profiling. This change does not alter our argument.

72. The frequentist argument in the trawl case is that the probability of interest is that of the event that the database search procedure will yield a match if indeed all of the individuals in the database are innocent. In the sequential search case, the analogy would be the probability that the sequential search procedure would yield a match if indeed all of the individuals sequentially investigated are innocent. This is effectively the classical statistical approach to sequential hypothesis testing.

73. See Balding & Donnelly, *Inference in Forensic Identification*, *supra* note 13, at 27-29.

DNA database matches the crime sample. Proof of a match between the crime sample and the profile of a given person is not weaker evidence, *with respect to the proposition that that person was the source of the crime sample*, because the person was identified as a suspect by the database search itself. And the fact that the search has eliminated some other persons as sources enhances the value of the evidence, albeit very slightly in the usual case.

D. *The Evidence as a Whole*

1. *Comparing the Confirmation and Trawl Cases*

Thus far, we have analyzed only the strength of the DNA evidence itself. But now consider the *entire body of evidence* in two cases, a confirmation case and a trawl case, in which the DNA profile of the crime sample is the same. For at least two reasons, the weight of the entire body of evidence may well be stronger in the confirmation case than in the trawl case.

First, by definition, in the confirmation case there is enough evidence independent of the DNA evidence to cast strong suspicion on the eventual defendant. By definition, that is not true before the database search in the trawl case, and it may not be true even after. And it may well be that, even after the trawl identifies the defendant as having DNA matching that of the crime sample, there is little or no other evidence tending to suggest that he is the perpetrator of the crime.⁷⁴ Assuming that this is true, and that the database searched in the trawl case was not very large in proportion to the suspect population, the entire body of evidence will plainly be stronger in the confirmation case than in the trawl case (though it may still be strong enough in the trawl case to warrant conviction). The prior odds of guilt are many times greater in the typical confirmation case than in the typical trawl case; if the likelihood ratio of the DNA evidence is approximately the same in the two cases, then the posterior odds — assessed after all the evidence is in — will also be much greater in the confirmation case than in the trawl case.⁷⁵

74. *Regina v. Adams*, [1998] 1 Crim. App. 377 (Eng. C.A. 1997), a rape case, is the textbook example. Effectively the only evidence linking the defendant to the crime was a DNA match between him and the crime sample, together with proof that he lived and worked in the general area where the rape took place. The victim, having originally told police that she had a good view of her attacker and that she remembered him clearly, did not identify Adams in a lineup. At the committal proceedings, she testified that he did not look like her description of her attacker; she stated that Adams looked to be in his forties while her attacker was in his early twenties. At trial, Adams testified that he had spent the whole of the night in question with his girlfriend, who corroborated his alibi.

75. Part C of the Appendix shows that, assuming that Matcher is just as likely, apart from the DNA evidence, as every other member of the population to be the source of the crime

Second, even if the evidence in the two cases, the non-DNA evidence as well as the DNA evidence, is comparable, it may appear to have greater weight in the confirmation case than in the trawl case. Suppose that in the confirmation case the police compile a powerful case against Matcher, based on circumstantial evidence and eyewitness identification, and only at the end of their investigation conduct the DNA test. And suppose that in the trawl case the police, having identified Matcher as a suspect only through a database search, focus their inquiries with such success that they are able to compile the same circumstantial and eyewitness identification evidence. In such a case, it may be plausible that the subsequently discovered evidence was tainted by suggestiveness, given that the DNA match motivated the police, and possibly witnesses as well, to confirm the suggestion that the perpetrator had been found.⁷⁶

These possibilities mean that the confirmation and trawl cases will not necessarily look the same, even assuming that the DNA evidence in the two is of a match with the same profile. But these are factors that a jury, aided by the arguments of counsel, can easily take into account. It is not hard, for example, for a defense attorney to argue, "Except for this DNA evidence, the prosecution does not have a shred of evidence against my client. And the prosecution's own expert acknowledged that there could well be several other people in the world with the same DNA." Similarly, defense counsel could argue, "The police and the eyewitnesses were hungry to find the perpetrator. After they got this DNA match, naturally they constructed a case to fit the hypothesis of my client's guilt." No technical expertise is necessary to make these arguments, and no adjustment in presentation of the DNA evidence is required. These factors may, however, account in part for the intuitive sense of some observers — including the NRC committees! — that the probative value of a DNA match is weakened by the fact that it is found after a database search. It is not that the DNA evidence is

scene sample, then the posterior odds that he is the source are R/N . Part C of the Appendix also shows that, for databases relatively small compared to the suspect population, these odds are much lower than the posterior odds in the confirmation case, even if the prior odds that Matcher is the source in that case are not very much greater than for other members of the population.

76. See Lempert, *Some Caveats*, *supra* note 11, at 319. Of course suggestiveness can cut both ways. In the confirmation case, the non-DNA evidence can create suggestiveness that might cast doubt both on subsequently discovered non-DNA evidence such as an eyewitness identification (if, for example, the police suggest that they have "got the man") and, given that the DNA test is performed by humans at the behest of the authorities, on the DNA evidence itself. We mean only to point out that the possibility of suggestiveness *might* weaken the force of the non-DNA evidence more in a given trawl case than in a comparable confirmation case.

weakened, we have argued; rather, it is possible that the *other* evidence in the case is weaker by virtue of the fact that identification from a database trawl has led to it.

2. *The Posterior Odds in the Trawl Case*

Now we will examine more closely the bottom line, the posterior odds — that is, the odds of the proposition at issue after all the evidence is considered — in the trawl case. Under Bayes's Theorem, as discussed above, these odds equal the prior odds times the likelihood ratio. Consider the highly schematized case in which all members of the suspect population are considered equally likely, apart from the DNA evidence, to be the source of the crime sample. Then, if testing of $D + 1$ samples yields exactly one match, that of *Matcher*, with that sample, and N persons in the suspect population remain untested, the posterior odds that *Matcher* is the source of the sample are

$$R/N. \qquad (3)$$

Ironically, the NRC likelihood analysis yields the same result.⁷⁷ This is because *in this situation* its two basic errors — focusing on the wrong hypothesis and failing to recognize the nature and weight of the DNA evidence — cancel each other out in determining the posterior odds.

Why is this so? The key proposition, we have said, is that *Matcher* is the source of the crime sample, and the evidence is that his sample, and none others of those tested, matched that sample. Under the NRC likelihood ratio approach, the key proposition is that the source of the crime sample is someone represented in the database, and the evidence is that exactly one such person (*Matcher*) had a sample matching the crime sample. *Given that Matcher, and nobody else tested, matched the crime sample*, the two propositions, that *Matcher* is the source and that someone in the database is the source, are identical. Thus, if the calculations are performed properly — and we do not suggest that the NRC or its supporters miscalculate — the posterior odds must be the same under the NRC's likelihood ratio analysis and under our own. In effect, what the NRC has done is to use too high a prior probability — because it uses an excessively broad hypothesis — and to com-

77. See Stockmarr, *supra* note 11, at 11. Stockmarr says the posterior odds are $k/(N - n)p$, where k is the number of matches, N the population, n the size of the database, and p the random match probability. For one match, this is identical to our R/N .

pensate by using a relatively small likelihood ratio that does not fully reflect the value of the evidence.

But if the bottom line, the posterior probability, is the same under either approach, does that mean that the debate is just a tempest in a teapot, an academic disputation with no real-world significance? Not at all. For one thing, we note parenthetically that it is *only* when the crime sample matches exactly one sample in the database, that of the eventual defendant, that the two propositions are identical. Of much greater practical importance, as we will now show, the NRC likelihood approach inevitably leads to flawed presentation of the evidence.

By definition, the prior odds have nothing to do with the DNA evidence; they are the odds that Matcher is the source of the crime sample as assessed apart from the DNA evidence. The expert presenting the DNA evidence therefore cannot, and should not, advise the jury on those odds. Presumably, in considering the non-DNA evidence, the jury is focusing on the ultimate issue before it:

(1) "How probable is it that *this defendant* committed the crime?"

And as part of their consideration of that issue, the jurors presumably consider the appropriate subsidiary issue:

(2) "How probable is it that *this defendant* was the source of the sample found at the crime scene?"

Presumably the jurors do *not* consider — because it is not part of the case — the question posed by the NRC:

(3) "How probable is it that *someone represented in the database* was the source of the sample found at the crime scene?"

What the expert can and should be advising the jury of is how the DNA evidence affects the odds on Question (2). In doing so, she should advise them of the likelihood ratio of the DNA evidence, or its components — that is, how probable it is that the DNA evidence would arise if Matcher were the source of the sample as compared to how probable it is that the evidence would arise if he were not. If instead the expert advises them of how the DNA evidence affects Question (3), she will be providing advice on an issue not before the jury. If, as appears likely, the jurors take the advice as bearing on Question (2) — because that is an issue that they *should* consider — then they will treat the evidence as far weaker than it really is.

In short, the expert's job is to advise on only half of the whole picture, and it is that half that the NRC likelihood analysis short-changes.

E. Presentation

We have touched above on, and will now consider more fully, the matter of presentation — that is, the issue of how the evidence should be presented to the jury in a way to convey full and accurate information and to minimize the chances of confusion.

We have assumed that it is appropriate in the confirmation case for the prosecution expert to report the match probability.⁷⁸ In the trawl case, the expert might testify in similar terms to the same probability; our analysis has shown that no adjustment to that probability is necessary. Thus, the expert might say:

The crime sample matches a sample that was taken from Matcher.⁷⁹
The probability that a sample taken from an arbitrarily chosen member of Defined Population would match the crime sample is 1 in R .

78. See *supra* section I.B.

79. This phrasing does not indicate when the sample was taken from Matcher. This is sometimes preferable, because revealing that the sample was taken before suspicion was cast on Matcher in this case, or that his profile was part of a database on file with the police, might suggest that he had a previous criminal record. Cf. *Commonwealth v. Claffey*, 400 A.2d 173 (Pa. Super. Ct. 1979) (emphasizing that jury did not learn that defendant was arrested after a fingerprint comparison proved positive, and thus had no reasonable basis for an inference of prior crime).

If the defendant's prior record is apparent to the jury anyway, there should be no problem. Cf. *State v. Lozano*, 588 P.2d 841, 842-43 (Ariz. Ct. App. 1978) ("Any prejudice was removed . . . when appellant testified in his defense and admitted a prior felony conviction that adequately explained the existence of the master [fingerprint] file card."); *People v. Garlin*, 428 N.E.2d 697, 702 (Ill. App. Ct. 1981). Also, the defendant may affirmatively wish the jury to know that the DNA identification was made before the police had any other significant evidence against him, so that he can argue that the other evidence was tainted by suggestion from the DNA identification. Cf. *State v. Montgomery*, 461 S.E.2d 732, 738-39 (N.C. 1995) (ruling that defendant opened the door to evidence concerning date of fingerprinting).

Even absent factors such as these, there is some reason to allow the prosecution to indicate, without flourishes gratuitously suggesting a prior record or its nature, that the defendant's profile was part of a database. This allows the prosecution to tell a coherent story of how the defendant's profile came to be compared to that of the crime scene sample. Furthermore, if the evidence that the other members of the database were eliminated as suspects is deemed significantly probative — which probably should be so only if the database is relatively large — it probably makes sense not to present the evidence as if the examination of the defendant's profile and the database trawl were two separate operations. Cf., e.g., *Garlin*, 428 N.E.2d at 701 (acknowledging that fingerprint card suggested prior arrest record, but holding that the evidence was necessary to prove identity).

The problem will probably be eased as DNA testing becomes more common, so that knowing that the police had access to a person's profile will not necessarily suggest that he had a prior record, or at least — as is already true in England — will not necessarily suggest that he had previously come under suspicion for a particularly serious crime. Cf., e.g., *Riley v. Sigler*, 320 F. Supp. 96, 99 (D. Neb. 1970) ("The mere fact that the jury knows that police officers have taken fingerprints of a defendant does not prejudice the defense. . . . [T]here was no assertion that the petitioner had previously been charged with a crime or as to why the fingerprint was on file. The jury could only speculate that petitioner had a record of other crimes." And the jury was told not to speculate. So there.); *People v. McGhee*, 605 N.E.2d 1039, 1048 (Ill. App. Ct. 1992) (holding, in a fingerprint case, that "the mere inference that defendant had committed some offense in the past, without some suggestion as to the nature and circumstances of that offense, fails to rise to a level of prejudice sufficient to overcome the obvious probative value such evidence provided in connecting defendant to this crime"); *Commonwealth v. Reiss*, 468 A.2d 451, 453 (Pa. 1983) ("Where there is no

But this statement does not convey the full extent of the DNA evidence in the trawl case; there is also the point that a number of possible suspects have been eliminated. Conveying this information raises a possibility of prejudice because it might suggest that the defendant was a contributor to a database, and this in turn might suggest that he had a prior criminal record.⁸⁰ If the probative value of this additional information is deemed to exceed its prejudicial impact, one way of conveying the information may be to allow the expert to add testimony in this form:

We also compared the crime sample to profiles from D other persons that we had on file, and none of those other profiles matched that of the crime sample.

F. *The Retesting Proposal*

As described in Section A, NRC I proposed that if a database trawl produced a match with the crime sample, this fact should only provide the basis for retesting the crime sample against a sample from the person so identified, using additional genetic markers not used in the original test. Thus, NRC I recommended that absent retesting the fact of a match discovered by a database search should not be admitted. Further, according to NRC I, assuming that the second test also indicated a match, only the markers used in that second test should be used in reporting the strength of the DNA evidence. Although NRC II criticizes the retesting requirement, such a requirement has recently been advocated by Professors Morton — who contends, indeed, that it is a sufficient response to the database search problem⁸¹ — and Lempert.⁸² For several reasons, we believe such a requirement is unwarranted.

indication . . . that the photographs were 'mugshots' or that they came from police files, it is not error for a witness to testify that he identified a defendant from photographs shown to him by police."); *State v. Emrick*, 282 A.2d 821, 826 (Vt. 1971) (refusing to find prejudice, because of common knowledge that fingerprints are taken for many purposes other than in connection with criminal proceedings).

80. See *supra* note 79.

81. See Morton, *supra* note 11, at 489. Morton agrees in principle with the analysis of NRC II. But he expresses concern that, for the defendant to gain the benefit of the diminution of probative value that he believes is appropriate, the defendant would have to reveal the fact that his profile was in a preexisting database, which may suggest a prior criminal record. This issue is addressed in note 79, *supra*. To avoid this problem — and others as well that are predicated on agreement with the fundamental analysis of NRC II — he advocates a retesting requirement.

82. See Lempert, *After the DNA Wars*, *supra* note 11, at 461. Lempert, like NRC II, believes that the probative value of evidence of a match found after a database trawl is diminished by the fact that, the larger the database, the more likely it is that a match will be found even though no one profiled in the database is actually the source of the crime sample. But then he criticizes the statistical adjustment offered by NRC II in an attempt to address this problem — for example, he discusses incisively the problem of the all-inclusive database

Perhaps most obviously, sometimes retesting is not feasible, even given due diligence on the part of the authorities.⁸³ It is possible to take further samples from a suspect, but the crime sample is fixed and not renewable. A test consumes part of the sample, and there may not be enough for a further test. This was a more serious problem with older testing techniques; even given the availability now of more sensitive techniques, however, it will still be a problem, albeit in a smaller and different set of cases.⁸⁴ Further, in some cases the sample may have become degraded over time or through mishandling that might be deemed excusable.

This concern about the feasibility of retesting might suggest not outright rejection of a retesting requirement but rather an exception to such a requirement. Under such a qualified requirement, evidence of a match found after a DNA search would be inadmissible without retesting unless (1) retesting is not feasible, and (2) this infeasibility is not attributable to the fault of the prosecution or its agents. But even such a qualified retesting requirement is inadvisable.

For one thing, a retesting requirement is not without costs. Obviously, even in the best of circumstances, there is the expense of the second test itself — and in other circumstances there is the cost of administering the rule, determining whether the particular case is one in which the absence of a second test is excusable. But the problem goes deeper. The prohibition against using the same markers in the two tests would make it difficult to achieve an optimal solution. One possibility — withholding from the first test good markers that would otherwise be used in it — is obviously unattractive because it diminishes the value of the first test for all searches. The other possibility — using an optimal number of markers in the first test and creating a backup set of additional

— and so concludes that this solution is “a step backward” from the retesting requirement of NRC I. (Professor Lempert was a member of the Committee that produced NRC I.)

83. See NRC II, *supra* note 10, at 134 (“If the amount of DNA in the evidence sample is too small, following the recommendation in the 1992 report could leave too few additional loci for computing a match probability or *LR*.”). The problem, where it exists, is substantially the same — insufficient additional DNA for a second test — whether the same or additional loci are used.

84. Any testing technique requires a minimum amount of DNA for it to work satisfactorily. This minimum amount is much smaller for more modern techniques than for older ones. But whatever the threshold may be for a given technique, there will be some cases in which the available DNA is barely above the threshold. In such cases, the first test will exhaust the DNA, leaving little or none for a second test. In other words, the increased sensitivity of modern tests shrinks the category of cases in which one test is possible but not two; as compared to previously, that category now includes cases involving smaller initial amounts of DNA. Even assuming that increased sensitivity shrinks that category to a smaller number of cases, it does not eliminate the category altogether.

markers for a second test sufficient to yield strong evidence on its own — creates a substantial burden on laboratories; it is expensive to develop satisfactory markers, and there is not a surfeit of approved markers available.⁸⁵

Some additional expense for the creation of substantially better evidence can be a worthwhile tradeoff, especially in a serious criminal case.⁸⁶ But courts should be wary of excluding DNA evidence on the grounds that better evidence would be possible. There is no limit to that logic, which would apply to all DNA testing, not only to identifications made after a database trawl. If using twelve markers is good, using thirteen is better — and fourteen better yet. But if the prosecution is satisfied to use a given piece of evidence, which is very probative and not plagued by defects that the defense cannot explain to the jury, the evidence should not be excluded on the grounds that the prosecution might have produced better evidence yet. The defense can, if it wishes, point out limitations of the evidence and opportunities the prosecution did not explore.

Perhaps most fundamentally, then, retesting in the trawl case should not be required because it is unnecessary. The DNA evidence in that case suffers no defect as compared to evidence of a match determined after a one-to-one comparison; indeed, as we have shown, the trawl result is somewhat better evidence. Excluding the proffered evidence in hopes of inducing the production of better evidence may be justifiable when there is some defect in the proffered evidence, or for some other reason the gap in quality between the proffered evidence and the other evidence is very great. But it is an altogether different, and unsatisfactory, matter to exclude the proffered evidence because its probative value has not been properly analyzed.

If a retest is conducted, and it confirms that the suspect's DNA matches that of the crime sample, the problem remains of what

85. Note, for example, the discussion of quality assurance and quality control in NRC II, *supra* note 10, at 76-78, (among other things endorsing the recommendations of NRC I, *supra* note 8, at 104-05, for thorough testing before implementation in casework). To gauge the extent of effort necessary to validate one particular collection of markers, see R. Sparkes et al., *The validation of a 7-locus multiplex STR test for use in forensic casework (I): Mixtures, ageing, degradation and species studies*, 109 INTL. J. LEGAL MED. 186 (1996); R. Sparkes et al., *The validation of a 7-locus multiplex STR test for use in forensic casework (II): Artefacts, casework studies and success rates*, 109 INTL. J. LEGAL MED. 195 (1996). In the words of Dr. Peter Gill, Head of Research at the United Kingdom Forensic Science Service and one of the authors of these papers, the development of a system of satisfactory markers involves "many person-years of work." Personal Communication with Peter Gill (Aug. 1998).

86. For a modern advocacy of the "best evidence" principle as the force behind much of evidentiary law, see Dale A. Nance, *The Best Evidence Principle*, 73 IOWA L. REV. 227 (1988).

should be reported to the jury. NRC I takes the view that only the results of the second test should be reported. But this solution discards perfectly useful evidence — the results of the first test — that the jury should be allowed to evaluate. Plainly, the fact that both the first and the second tests produced a match between the suspect and the crime sample — and that none of the other profiles reviewed in the first test did so — is more powerful evidence than proof of the second test alone.

III. SOURCES OF DIFFICULTY: STATISTICAL ASPIRATIONS TO SCIENCE AND UNDUE JUDICIAL DEFERENCE

In Part II, we have contended, contrary to the NRC reports and some distinguished scholars who have supported them, that the probative force of a DNA match is not weakened by the fact that it was made only by virtue of a database trawl. If we are right, and in this Part we will assume that we are, then the question arises of how such a flagrant error has been made. We have already suggested one pair of reasons — that the probative value of the other evidence in the case may be weakened by the fact that the DNA match has been discovered through a trawl, and that the total weight of the evidence may be less in a trawl case than in a comparable confirmation case. In this Part, we suggest other, deeper reasons. We explore both the habits of mind of statisticians that make them prone to errors of the type we have discussed here, and a judicial tendency towards deference that diminishes the ability of courts to make good use of statistical and scientific evidence.

A. *Classical Statistics and the Aspiration to Objectivity*

1. *The Classical Statistical Model*

Science aspires to objectivity, to the demonstration of propositions that are not dependent on the subjective views of the observer.⁸⁷ Accordingly, it is highly dependent on experiments in which given sets of conditions are observed many times. By counting or measuring different consequences in different conditions, a scientist can hope to draw conclusions on the associations between conditions and consequences.

87. See, e.g., Larry Laudan, *Why do (natural) scientists agree?*, in *INTERPRETING THE WORLD* 89 (William R. Shea & Antonio Spadafora eds., 1992). Although controversy remains among philosophers of science over the extent to which such objectivity is attainable, we do not intend to address this issue here. For our purposes, the general aspiration of science towards objectivity is the critical point.

Accordingly, classical statistics grew up to facilitate objective inferences from data. Classical statisticians try to avoid subjective judgments, seeking instead to determine what conclusions can be drawn solely on the basis of frequency of observation. The Bayesian approach — updating the odds assigned to a given proposition in light of evidence subsequently received — is thus unacceptable to classical statisticians because it depends on the subjective assignment of odds in the absence of objectively measurable data. Instead, the classical statistician, having selected a hypothesis to be tested but without having assigned any probability to that proposition, observes whether the results of an experiment are of a highly unusual nature assuming the truth of that proposition. If they are, then the statistician concludes that the hypothesis can be rejected. Often the statistician works by indirection; the “null hypothesis,” the hypothesis to be tested and perhaps rejected, is the negation of the hypothesis the statistician is hoping to confirm.

Under this approach, great caution must be exercised in testing several different hypotheses in the same experiment. The reason is that the more different kinds of observations are made, the more likely it is that some of the observations will be unusual in nature, even though in fact nothing remarkable has happened. Take an illustration adapted from one offered by NRC II. Suppose we have some reason to hypothesize that a coin is biased towards heads, and to test this we toss it in the air five times, with all five tosses landing heads up. This provides some basis to reject the hypothesis that the coin is fair, because this result would occur in only about three percent of all repetitions of the experiment with a fair coin.⁸⁸ But now suppose we hypothesize that each of ten coins might be biased towards heads, and we toss each of the ten five times. If one or more of the coins lands heads up on all five tosses, the classical statistician maintains that this provides no good reason for concluding that such coins are biased; in approximately twenty-seven percent of all repetitions of this experiment with ten fair coins, at least one coin will land heads up on all tosses.⁸⁹ Given 100 coins, it is virtually certain that at least one will land heads up on all five tosses — even though all 100 are perfectly fair.

88. More precisely, in $1/2^5$, or $1/32$, of all repetitions.

89. The probability that a given fair coin will land heads up on all 5 tosses is $1/32$, so the probability that it will not is $31/32$. Accordingly, the probability that none of 10 fair coins tossed independently will land heads up on all 5 tosses is $[31/32]^{10}$, which equals approximately 73%.

For the same reason, the classical statistician is very reluctant to draw proof of a proposition from a set of observations unless the observations have been made with the proposition (or its negation) articulated beforehand as a hypothesis to be tested. Given enough observations, it is highly likely that some unusual ones will turn up, even though no surprising causal factor is at work; the one-in-100 longshot comes in a winner about one time in 100, and when it wins that does not mean that anything suspicious has happened. It is considered bad science to trawl data, looking for surprising results and then proclaiming that the data proves a proposition that would likely lead to such results. At best, such unusual results help identify a hypothesis to be tested. New observations, perhaps from a new experiment, are then necessary to test the truth of that hypothesis.

We will not attempt here to evaluate how well this classical statistical model performs the job for which it was designed, aiding the interpretation of data in scientific inquiry; that is a broad and controversial issue beyond the scope of this Article.⁹⁰ Rather, we will argue that the NRC approach to the database search problem reflects the classical statistical method and that this is an inappropriate method for resolving such problems in the distinctive setting created by adjudication.

2. *The NRC's Classical Orientation*

That the NRC approach reflects the classical method is rather obvious on the face of the reports. NRC I emphasized that with a database trawl "one does not start with a single hypothesis to test (i.e., that the evidence was left by a particular suspect), but instead fishes through the databank, trying out many hypotheses."⁹¹ Thus, no conclusion could be drawn from the fact of a match produced by the trawl; instead, the match "should be used only as the basis for further testing using markers at additional loci."⁹² In other words, the initial match yielded by the search identifies the hypothesis to be tested, and nothing more. Though NRC II did not recommend the same solution, its analysis of the problem was very similar. Its reference to the teachings of "standard statistical logic" in a coin-toss experiment is revealing: In one experiment, a heads up toss of each of twenty coins is so unusual that it supports rejection of the

90. See generally JOHN EARMAN, *BAYES OR BUST* (1992); GERD GIGERENZER ET AL., *THE EMPIRE OF CHANCE* (1989).

91. NRC I, *supra* note 8, at 124.

92. See *id.* at 129; *supra* text accompanying note 38.

null hypothesis that the coins are all unbiased. But if the experiment is repeated a large number of times, an all-heads-up result on at least one trial proves nothing because given enough trials it is unsurprising that some of them will have unusual results. And this analogy, NRC II maintained, was on point for the database search problem, because the more profiles searched the greater the chance of finding one that matches the crime sample purely by coincidence.

3. *Conflict with the Needs of the Legal System*

The enterprise of law is very different in several ways from the scientific enterprise for which the classical statistical model was developed.

a. Nonrecurrence and The Need for Decision. In serving science, the classical statistical method seeks to ascertain recurrent patterns, using observations about occurrences in the past to predict the future. The future is infinite, and so science is, in the main, patient; a scientist following the dictates of classical statistical methods will not draw a conclusion until she is confident she can do so.

Adjudication, by contrast, usually seeks to determine the truth about a particular dispute, which usually concerns events in the past. The job of a jury — which we will use as a shorthand for adjudicative factfinders — is not to draw general conclusions about a run of cases, but rather to determine the truth of the dispute before it. It must do so, as best it can, on the basis of the information legitimately before it; it cannot decline to decide because it does not have sufficient information. The inertial decision — leaving matters where they stand — does not simply leave matters open for resolution another day. Rather, it is a final decision against the claimant who sought to alter the status quo. And in making its decision, the jury can rely to a limited extent, if at all, on information about other events, because the case before it is a non-recurrent event that is almost certainly unique in some material respect.

b. Subjectivity. In part because of the non-recurrence of litigated disputes, adjudicative fact-finding does not usually purport to depend on objective assessment of quantitative data. Most of the evidence presented to juries is not in quantified form. Hence, the legal system expects juries to make subjective assessments of the evidence. It recognizes that different juries might rationally make substantially different assessments of a given body of evidence.⁹³

93. Thus, although an American court in a civil case has the power to throw out a verdict as against the great weight of the evidence, if the case goes to the jury then a verdict for either side will usually be upheld. See, e.g., *Wilkerson v. McCarthy*, 336 U.S. 53, 55 (1949)

Thus, the legal system embraces subjectivity, for it has no other choice. And this means that, while the system does not usually have the opportunity of relying on quantified data, it is not shackled by the necessity of doing so. In other words, a jury is expected to make its subjective assessment of the probability of a proposition at issue on the basis of all the information legitimately available to it. Differences in this base of information will therefore reasonably lead to different assessments of such probabilities. Indeed, the standard definition of relevant evidence is "evidence having any tendency to make the existence of any fact that is of consequence to the determination of the action more probable or less probable than it would be without the evidence."⁹⁴

The need for subjectivity applies to investigators, or inquisitors, as well as to jurors. Near the beginning of the investigation, when he may have no primary suspect in mind, the investigator must in effect treat as a suspect everybody in the world capable of committing the crime. He must, however, make subjective assessments — some persons will be far more likely suspects at the outset than others — and must continually adjust those assessments as he receives new information. If the investigator learns information strongly pointing to one person, the proposition that that person was the perpetrator immediately becomes of interest, even if the prior probability of that proposition appeared very low.

c. Pre-Set Issues and Information. A jury is relatively passive, in two interrelated senses. First, to a substantial extent the propositions at issue are identified before the case is presented to the jury. In a criminal case, a particular crime or set of crimes is at issue, and a particular defendant or set of defendants is on trial. Thus, in a one-defendant trial, if the jurors are persuaded beyond a reasonable doubt that the alleged crime was committed, their question is not, "Who did it?" but rather, "Did the defendant do it?"⁹⁵ Second, for the most part, the jury cannot demand the production of more evidence to test hypotheses that it finds of interest; it must do the best it can with the evidence presented to it and the knowledge

("This Court has previously held in many cases that where jury trials are required, courts must submit the issues of negligence to a jury if evidence might justify a finding either way on those issues.").

94. FED. R. EVID. 401.

95. The presumption of innocence demands that, before the presentation of evidence, the jury assign a very low probability to the proposition that the defendant was the perpetrator; in particular, the jury must not allow the fact that the prosecution has chosen to bring charges against the defendant, alone among all other people on earth, to elevate that probability.

it is allowed to bring into the deliberation room. A jury has very limited capacity to experiment.

This latter point applies, though perhaps with less force, to investigators (or inquisitors) as well. An investigator does have some ability to select a hypothesis and then search for evidence that will tend to confirm or disprove it, and sometimes he can even create an experiment that will do so. But in most cases the investigator inquires about events in the past. This means that for the most part his job is not to create new information, but to discover, assemble, and interpret the information that is available.

d. Summary. The classical statistical approach is unsuited for analyzing DNA evidence in the context of litigation. The problem facing an investigator or a juror is not to determine a general law of the universe. Rather, the investigator must try to determine who committed a given crime on a given occasion; the juror's job is even narrower, seeking to determine whether a given person, on whom its attention has been focused, committed the crime. Subjective assessment, in light of all the information legitimately before them, is essential in allowing both the investigator and the juror to perform their jobs properly. Sufficiently strong evidence may reasonably lead to the conclusion that a given person (whether or not previously identified as a particular source of suspicion) is the perpetrator even though the prior probability of that proposition seemed very low. Once the case is presented to the jury, the jury cannot decide it lacks the information to draw a conclusion, nor that it wants to perform an experiment to ascertain the facts: It must make its best assessment of the facts on the basis of the information within its purview. Thus, where the evidence tends to show that the person who left a particular DNA sample at a particular place is probably the perpetrator, it becomes of importance to the jury whether the defendant is the source of that sample. The jury should compare how likely the DNA evidence would arise if the defendant were the source to how likely it would arise if he were not the source. And the jury should combine these assessments with its subjective assessments of all the other evidence in the case to assess the probability of guilt.

B. *The Law as Consumer of Forensic Science*

We have, we think, shown why statisticians are prone in a trawl case to provide the adjudicative system with the answer to the wrong question. But why does the legal system appear to be ready to accept that advice? The answer, we believe, lies in a longstand-

ing tendency of the adjudicative system to defer to the scientific establishment.

Courts have long been afraid that juries will fall prey to "junk science."⁹⁶ Thus, they have demanded that an expert witness's opinion be in accordance with a theory that has achieved some threshold level of reputability. For most of this century, the dominant statement of this idea was the one in *Frye v. United States*, that the underlying "scientific principle or discovery . . . from which the deduction is made must be sufficiently established to have gained general acceptance in the particular field in which it belongs."⁹⁷

Some states still adhere to the *Frye* test.⁹⁸ But many jurisdictions have justifiably come to the conclusion that its demand for "general acceptance" by the relevant scientific community as a precondition to admissibility is too stringent. Federal Rule of Evidence 702, which has been adopted in many states, supports this view. That Rule provides:

If scientific, technical, or other specialized knowledge will assist the trier of fact to understand the evidence or to determine a fact in issue, a witness qualified as an expert by knowledge, skill, experience, training, or education, may testify thereto in the form of an opinion or otherwise.⁹⁹

In 1993 in *Daubert v. Merrell Dow Pharmaceuticals, Inc.*,¹⁰⁰ the United States Supreme Court held unanimously that the "austere standard" of *Frye* is incompatible not only with the language of Rule 702 but also with the liberal nature of the Rules in general.¹⁰¹ A majority of the Court went beyond this ruling, however, and attempted to articulate the "gatekeeping role" that the judge must play when "[f]aced with a proffer of expert scientific testimony."¹⁰²

Justice Blackmun's opinion for the majority expressly limited this discussion to evidence based on scientific knowledge, as opposed to "technical, or other specialized knowledge," because that was the nature of the expertise offered in *Daubert* itself.¹⁰³ Justice Blackmun put great emphasis on Rule 702's use of the term "scien-

96. See, e.g., PETER HUBER, *GALILEO'S REVENGE: JUNK SCIENCE IN THE COURTROOM* (1991).

97. 293 F. 1013, 1014 (D.C. Cir. 1923).

98. See, e.g., *People v. Leahy*, 882 P.2d 321 (Cal. 1994) (declining to follow *Daubert* and continuing to apply California's version of *Frye*, commonly referred to as *Kelly-Frye*); *People v. Peterson*, 537 N.W.2d 857 (Mich. 1995).

99. FED. R. EVID. 702.

100. 509 U.S. 579 (1993).

101. 509 U.S. at 589.

102. 509 U.S. at 597, 592.

103. 509 U.S. at 590 & n.8.

tific . . . knowledge” and operated from the premise that, to qualify as scientific knowledge, “an inference or assertion must be derived by the scientific method.”¹⁰⁴ Thus, he attempted to articulate indicia of the scientific method, and he laid out four criteria that should often enter into determining “the scientific validity” of the principles underlying the evidence. These may be referred to as (1) testing, (2) peer review and publication, (3) error rates and standards, and — partially resurrecting *Frye* just a few pages after its apparent death — (4) general acceptance.¹⁰⁵

Our concern here is not so much with the much debated questions of whether *Daubert* reflects good philosophy of science, or even whether it reflects good evidentiary policy.¹⁰⁶ Rather, the point of significance here is that, though *Daubert* reflects a loosening of the demands purportedly applied under *Frye*, it still reflects a notable attitude of deference to the scientific establishment. This attitude is apparent in several respects. Most obviously, perhaps, is the continued use of “general acceptance” as a criterion — albeit no longer the exclusive one — for determining admissibility. Further, the other criteria constitute an adoption of currently prevalent scientific methods; the emphasis on peer review and publication also relies heavily on the attitudes of, and decisions made by, the scientific establishment. Perhaps most fundamentally, the entire inquiry seems gratuitous. Under the language of Rule 702, nothing seems to depend on whether the knowledge on which the opinion is based is “scientific” or not. “[S]cientific” and “technical” are clearly listed merely as illustrations of “specialized knowledge”; the key question seems to be whether the opinion is based on “specialized knowledge [that] will assist the trier of fact to understand the evidence or to determine a fact in issue.”¹⁰⁷

The bite of *Daubert*'s insistence on scientific methods was demonstrated recently in *General Electric Co. v. Joiner*.¹⁰⁸ There, the plaintiff attempted to show that exposure to PCBs, a class of chemicals, could promote cancer. He presented the results of four

104. 509 U.S. at 590.

105. 509 U.S. at 592-95.

106. See, e.g., Heidi Li Feldman, *Science and Uncertainty in Mass Exposure Litigation*, 74 TEXAS L. REV. 1 (1995); Brian Leiter, *The Epistemology of Admissibility: Why Even Good Philosophy of Science Would Not Make for Good Philosophy of Evidence*, 1997 B.Y.U. L. REV. 803 (1997).

107. FED. R. EVID. 702. Justice Blackmun did purport to focus on the “helpfulness” standard — but he did so by melding that issue into the one of scientific validity. See *Daubert*, 509 U.S. at 591-92 (“Rule 702’s ‘helpfulness’ standard requires a valid scientific connection to the pertinent inquiry as a precondition to admissibility.”).

108. 118 S. Ct. 512 (1997).

studies, each of which failed on its own to provide strong support for this conclusion. One of his experts, a toxicologist named David Teitelbaum, testified — somewhat ungrammatically, but comprehensibly enough — at a deposition:

[A]s a toxicologist when I look at a study, I am going to require that that study meet the general criteria for methodology and statistical analysis, but that when all of that data is collected and you ask me as a patient, “Doctor, have I got a risk of getting cancer from this?” That those studies don’t answer the question, that I have to put them all together in my mind and look at them in relation to everything I know about the substance and everything I know about the exposure and come to a conclusion. I think when I say, “To a reasonable medical probability as a medical toxicologist, this substance was a contributing cause,” . . . to his cancer, that that is a valid conclusion based on the totality of the evidence presented to me. And I think that that is an appropriate thing for a toxicologist to do, and it has been the basis of diagnosis for several hundred years, anyway.¹⁰⁹

Justice Stevens agreed with Dr. Teitelbaum that “[i]t is not intrinsically ‘unscientific’ for experienced professionals to arrive at a conclusion by weighing all available scientific evidence — this is not the sort of ‘junk science’ with which *Daubert* was concerned.”¹¹⁰ But Justice Stevens stood alone. The rest of the Court upheld the trial court’s decision that expert opinions like this one did not rise above “subjective belief or unsupported speculation.”¹¹¹ A trial court, concluded the majority, could validly decide to exclude “evidence which is connected to existing data only by the *ipse dixit* of the expert. A court may conclude that there is simply too great an analytical gap between the data and the opinion proffered.”¹¹²

The fear of junk science — the concern that juries will be overwhelmed by chicanery masquerading as science — has thus exerted a powerful effect on the American courts. And perhaps a fear of their own inadequacy to separate the wheat from the chaff has led them to rely greatly on the attitudes and methods of the scientific establishment. We do not mean to deny that a “gatekeeping role” for the courts is necessary; we assume that some evidence offered under the guise of science is of so little value, and of sufficient prejudicial potential, that exclusion is warranted. But we believe that deference to the scientific establishment in an attempt to fend off junk science may create another problem — failure to recognize the

109. 118 S. Ct. at 522 n.4 (quoting David Teitelbaum).

110. 118 S. Ct. at 522 (Stevens, J., dissenting).

111. 118 S. Ct. at 516.

112. 118 S. Ct. at 519.

extent to which, as we have suggested in Section A of this Part, the needs of the legal system do not match up with the methods ordinarily used by scientists.

The *Joiner* majority's castigation of the toxicologist's opinion as "subjective belief" reflects this failure. Adjudication, we have argued, depends on subjective assessments of factual issues, and courts are fooling themselves if they pretend otherwise.¹¹³ Sometimes, scientists can present to the jury generalized propositions of the type that they try to demonstrate in their ordinary, non-forensic work. But adjudication usually depends on the particulars of the case at hand.

Often this means that the law needs to decide non-recurrent matters for which, because it is impossible to run a controlled experiment or even to gather data across like cases, the scientific method will be useless. And yet, in such cases scientifically-based information may be useful in trying to determine the facts. The subjective belief of an expert who has had extensive experience in dealing with problems of a roughly similar nature may be particularly useful in bridging the gap between those principles and the available evidence. This is often the case, for example, in cases in which an engineer offers an explanation for an accident that is similar in some respects to others but unique in other respects.¹¹⁴ Even if the expert's opinion is not well grounded on scientific principles, her observations and judgment, based on extensive experience, may be useful.¹¹⁵ For this reason, we agree with the orientation of decisions like *Carmichael v. Samyang Tire, Inc.*,¹¹⁶ now on review before the Supreme Court, which treat the *Daubert* criteria as limited to the class of expert opinion the admissibility of which depends on its being deemed to be scientific.

Even if the matter on which an expert wishes to offer an opinion is a recurrent one, so that science can in time yield an answer with

113. See, e.g., EVETT & WEIR, *supra* note 13, at 240-41.

114. But see, e.g., Peitzmeier v. Hennessy Indus., Inc., 97 F.3d 293, 297 (8th Cir. 1996) (applying *Daubert* criteria, over plaintiffs' protests that the expert's testimony was to matters of basic engineering), cert. denied, 117 S. Ct. 1552 (1997). Cf. Gier v. Educational Serv. Unit No. 16, 66 F.3d 940, 943-44 (8th Cir. 1995) (applying *Daubert* to psychological evaluations in child custody cases).

115. See, e.g., United States v. Starzecpyzel, 880 F. Supp. 1027 (S.D.N.Y. 1995) (holding that the work of forensic document examiners is not genuine science, but nevertheless admitting an opinion as a non-scientific one in a case of forgery detection because it could help the jury in focusing on minute similarities and dissimilarities between two documents that laypeople might have missed).

116. 131 F.3d 1433 (11th Cir. 1997), cert. granted *sub nom.* Kumho Tire Co., Ltd. v. Carmichael, 118 S. Ct. 2339 (1998).

confidence, science may not be ready to do so before the legal system needs guidance. Unlike scientific inquiry, Justice Blackmun pointed out in *Daubert*, law “must resolve disputes finally and quickly”; evidentiary rules are “designed not for the exhaustive search for cosmic understanding but for the particularized resolution of legal disputes.”¹¹⁷ But the Court seems to have failed to realize the implications of that perspective. It does not weigh in favor of excluding scientific evidence, or of deferring to the scientific establishment, but rather in favor of recognizing that the law must satisfy its own needs for scientific advice, even if doing so does not square with the usual methods of scientists.¹¹⁸

In any case, if a scientist is doing more than reciting general principles without an attempt to relate them to the facts of the case, the law’s treatment of scientific evidence must take into account the adjudicative context in which the evidence is offered. That context differs significantly from the one in which scientists are used to working — most notably, with respect to evidence of a DNA database search, in that the jury’s job is at base a subjective one and in that the bringing of the case tends to define the propositions at issue.

Our perspective may perhaps be crystallized by comparing it to that of Anders Stockmarr, one of the statisticians who has written in support of the NRC approach to the DNA database search problem. “The decision problem of the court,” he has written, “should take the implications of statistical hypotheses for data description into account, and not the other way around.”¹¹⁹ In our view, this is precisely wrong. The legal system is a consumer of the information offered by expert witnesses. It may be that the service needed by the legal system requires scientists to operate in ways at variance with their usual operating methods. The law should not be a passive consumer of scientifically based information, taking what scientists have to offer “off the rack.” Rather, it should be an aggressive consumer, asking its suppliers to provide what it needs.¹²⁰

117. *Daubert v. Merrell-Dow Pharmaceuticals*, 509 U.S. 579, 597 (1993).

118. See Richard D. Friedman, *The Death and Transfiguration of Frye*, 34 JURIMETRICS J. 133, 147 (1994) (“[S]ometimes courts ought to be willing to allow juries to take advantage of scientific information even when the scientific establishment is unwilling to declare a conclusion.”).

119. Stockmarr, *supra* note 11, at 15.

120. See generally Leiter, *supra* note 106, at 805, 814 (arguing that the questions “What is the best account of scientific method?” and “What is the best criterion for judges to use in deciding the admissibility of scientific evidence?” must be kept separate, and posing the question of “social epistemology:” “[U]nder the real-world epistemic limits of a particular

This perspective, which has sometimes been apparent in debates concerning psychiatric testimony,¹²¹ may be helpful across the range of expert testimony.¹²² What the law needs is not necessarily

social process for the acquisition of knowledge, what epistemic norms actually work the best?”).

121. The law sometimes attaches serious consequences to the matter of an individual's mental condition or constitution. Where those criteria do not accord well with criteria ordinarily used by psychiatrists, or where the law asks psychiatrists for predictions that the profession believes cannot be made with confidence, many psychiatrists naturally chafe. See, e.g., *Barefoot v. Estelle*, 463 U.S. 880 (1983) (over dissent, rejecting position of American Psychiatric Association that psychiatric predictions of long-term future dangerousness are too unreliable to be admissible); Statement of Dr. Lawrence C. Kolb, Director, N.Y.S. Psychiatric Institute, at Second Circuit Conference, 37 F.R.D. 365, 387 (1964) (“[M]any of our younger people have very little interest today in offering themselves as experts in court in view of the constricted framework in which they believe their testimony may be cast, a framework which is considered irrelevant to their areas of competence.”). At times, the response of the courts in the context of the insanity defense has been to defer, allowing testimony in terms of criteria approved by the psychiatric profession. See, e.g., 2 PAUL H. ROBINSON, *CRIMINAL LAW DEFENSES* § 173(b)(3) (1984) (noting and criticizing “a common tendency to define the mental illness disability in terms of psychiatric labels”).

We do not wish to minimize the difficulties; if, for example, predictions of violence cannot be made to an acceptable degree of confidence, then it is probably unwise to create a legal rule that depends on such predictions. But determining what constitutes sufficient mental capacity for a person to be held responsible for a crime is ultimately a normative judgment. See, e.g., *United States v. Brawner*, 471 F.2d 969, 983, 981 (D.C. Cir. 1972) (en banc) (rejecting the *Durham* rule — which asked whether the crime was the “product” of a mental disease or defect and “was adopted in large part to permit experts to testify in their own terms” — principally because the rule was responsible for the “undesirable characteristic . . . of undue dominance by the experts giving testimony”); *McDonald v. United States*, 312 F.2d 847, 851 (D.C. Cir. 1962) (en banc) (“[N]either the court nor the jury is bound by *ad hoc* definitions or conclusions as to what experts state is a disease or defect. What psychiatrists may consider a ‘mental disease or defect’ for clinical purposes, where their concern is treatment, may or may not be the same as mental disease or defect for the jury’s purpose in determining criminal responsibility.”); 312 F.2d at 861 (Miller, C.J., concurring in part) (“Until now, this court has allowed the shifting wind of expert nomenclature to control its decisions.”); ROBINSON, *supra*, § 173(b)(3) (emphasizing “the community’s view” as “central to the proper operation of a system of excuses”). Psychiatrists can be enormously helpful, both in advising on the practicality and consequences of any test and in providing information and opinions to help apply it. At bottom, though, crafting the test is a matter of law.

122. For example, scientific experts sometimes testify that a given result is “consistent” with a proposition at issue. See, e.g., ROBERTSON & VIGNAUX, *supra* note 20, at 56. This tendency may well be a result of a phenomenon on which we have commented, the hesitancy of scientists to reach definitive conclusions. It may also reflect a habit of thinking consonant with classical statistics, in which an analyst seeks to determine the validity of a proposition by subjecting it to one or more tests, rejecting the proposition if it is inconsistent with the test results. In any event, this manner of testifying is unhelpful for the law. An adjudicative factfinder needs to know the relative likelihood of the result given the truth of the proposition at issue and given its falsity. This information is obscured by testimony in terms of consistency.

We offer one more example from the realm of statistics. Meta-analysis is the use of formal statistical techniques to provide a quantitative synthesis of multiple studies, each of which may in itself be insufficient to yield significant results. The use of meta-analysis has been controversial in various disciplines, see, e.g., Jack W. Snyder, *Silicone Breast Implants*, 18 J. LEG. MED. 133, 209 (1997). We do not wish here to suggest criteria for the use of meta-analysis in adjudication, or to minimize the difficulties of drawing conclusions from a body of studies that may use somewhat different methodologies. But our model of the law as an aggressive consumer suggests that courts should probably be willing — under the necessity of drawing the best conclusion possible given the available data — to allow inferences to be

information processed in the usual ways of science, but rather information that will be helpful to the jurors in making their best subjective assessment of the particular issues at stake in the case at hand. Of course, like any consumer, the law can only ask the supplier to provide what the supplier can. But in the implicit negotiation between law and science, the law has one advantage that most consumers do not have: It sets the rules.¹²³

CONCLUSION

We have argued here that the NRC, and those who have supported it, have erred in denigrating the value of evidence of a match found after trawl of a DNA database. This is an issue likely to be of great importance with the continued growth of offender databases.

One root of the problem, we believe, is the classical statistical orientation of many statisticians. Because of their fear of testing multiple hypotheses, a fear based on the use of statistics in a supposedly objective manner for scientific inquiry, they focus on the wrong question, asking how the DNA evidence alters the probability of the proposition that the source of the crime sample is profiled in the database. Instead, they should ask how the DNA evidence alters the probability of the proposition that the person identified by the trawl was the source. And the adjudicative system seems ready to accept this mistaken analysis because of a habit of deference to the scientific establishment borne out of a fear of junk science. If the courts would recognize that what they need from science is not the usual output of the scientific community, but rather a special product more tailored to adjudicative needs, they may better play the role of aggressive consumer, and so better secure information that will be of help to the trier of fact.

drawn from meta-analyses under criteria that in non-forensic contexts would appear unduly lax.

123. See generally Peter H. Schuck, *Multi-Culturalism Redux: Science, Law, and Politics*, 11 YALE L. & POL. REV. 1 (1993). Schuck "advocate[s] a criterion of cultural competence for allocating decisional authority over multi-cultural issues," *id.* at 3, and cautions:

Scientists must remember that it is lawyers and politicians who formulate the public rules for our complex society. Lawyers must remember that the project of science demands, and generally deserves, a kind of freedom to which legal controls will often be inimical. Politicians must remember that science . . . is to some irreducible degree an elite enterprise that cannot flourish under the incubus of a militant populism.

Id. at 46.

APPENDIX

To highlight some of the points made in the text, this Appendix presents an informal mathematical discussion of a simplified model of inferential problems arising from DNA evidence.¹²⁴

Assume that in both the confirmation and trawl cases the crime sample and the sample taken from Matcher have the same DNA profile, and that the probability that any person arbitrarily chosen from the suspect population will have that profile is p .¹²⁵ We will disregard complexities created by relationships among members of the population, and the possibility of error in handling or analyzing the evidence.

A. *The Likelihood Ratio in the Confirmation Case*

In the confirmation case, the forensic scientist could report the match probability p . Another way of phrasing it, in numbers greater than rather than less than one, is to say that the random match probability is "1 in R ," where R equals $1/p$. The likelihood ratio is also very simple. The numerator is the probability that the evidence of a match would arise if Matcher were the source of the crime sample. That probability, under the simplifying assumptions we have made, is 1. The denominator is the probability that the evidence of a match would arise if Matcher were *not* the source of the crime sample. That probability, under the simplifying assumptions we have made, is p . The likelihood ratio is therefore $1/p$, or R .

B. *The Likelihood Ratio in the Trawl Case*

Now consider the trawl case. There is a database consisting of samples taken from $D + 1$ persons, and in addition to those people anyone within a suspect population of N people might have been the source of the crime sample. Assume that, before the database is trawled, all people represented in the database are deemed equally likely to be the source of the crime sample, that all members of the suspect population not represented in the database are deemed equally likely to be the source, and that any given person represented in the database is deemed m times more likely than any

124. For analyses of more complex and general versions of the problem, see Balding & Donnelly, *Evaluating DNA Profile Evidence*, *supra* note 13, and Balding & Donnelly, *Inference in Forensic Identification*, *supra* note 13.

125. In reality, learning information about the profiles of some members of the population gives new information bearing on the probability that any other member of the population will have the profile in question. We will disregard this complexity for purposes of this discussion.

member of the population not represented in the database to be the source. The database is then trawled, and only one sample, that of Matcher, matches the profile of the crime sample. We can now calculate the likelihood ratio for this case.

The numerator, again, is the probability that the evidence would arise if Matcher were the source of the crime sample. But now the evidence is not simply that Matcher's profile matches that of the crime sample; it is also that the other D samples do not match the crime sample. The probability that Matcher's profile would match that of the crime sample, given that he is the source of that sample, is again 1. The probability that any given profile of an innocent person in the database would not match that sample by chance is $(1 - p)$. Under the simplifying assumptions we have made, the probability that *none* of the D profiles in the database other than Matcher's would match the crime sample if Matcher is the source of the sample is therefore

$$(1 - p)^D. \tag{A1}$$

Given the certainty that Matcher's sample would match that of the crime sample if he is indeed the source of that sample, this expression is the numerator of the likelihood ratio.

Now we must determine the denominator of the likelihood ratio, which is the probability that the evidence would arise if Matcher were *not* the source of the crime sample. This is somewhat more complicated because we must take into account two basic possibilities — (1) that the source was someone other than Matcher represented in the database, and (2) that the source was a member of the population not represented in the database. Before the evidence of the trawl is received, the first of these possibilities appears mD/N times as likely as the second. These two possibilities must be weighted in this proportion in determining the probability that the DNA evidence would arise if Matcher were not the source. That probability is therefore

$$\frac{mD \times Pr(E|S_d) + N \times Pr(E|S_p)}{mD + N}, \tag{A2}$$

where $Pr(E|S_d)$ is the probability that the evidence would arise if someone other than Matcher represented in the database were the source of the crime sample, and $Pr(E|S_p)$ is the probability that the evidence would arise if someone in the suspect population but not represented in the database were the source.¹²⁶

126. In general, if H equals $(H_1 \text{ or } H_2)$, and H_1 and H_2 are mutually incompatible,

$Pr(E|S_d)$ is easy to determine. It is 0, for if the source were someone represented in the database other than Matcher, it would be impossible for the evidence — that Matcher and no one else in the database matched that sample — to arise.

$Pr(E|S_p)$ measures the probability that, if someone not represented in the database were the source of the crime sample, (1) Matcher's profile would match that of the crime sample by chance and (2) none of the other profiles D in the database would do so. Because we are assuming that these two events are independent, the probability of their conjunction is the product of the individual probabilities. The first of these individual probabilities is p , the random match probability. The second is $(1 - p)^D$, the same as (A1). Thus, $Pr(E|S_p)$ equals

$$p(1 - p)^D. \tag{A3}$$

From (A2), the denominator of the likelihood ratio therefore equals

$$\frac{0 + N \times p(1 - p)^D}{mD + N}. \tag{A4}$$

The likelihood ratio equals (A1) divided by (A4), which simplifies by elementary algebra to

$$1/p \times (1 + mD/N). \tag{A5}$$

Because R is defined to be $1/p$, this simplifies further to

$$R \times (1 + mD/N), \tag{A6}$$

which is Expression (2) in the text. Expression (A6) has a simple interpretation: It may be rewritten as $R \times (N + mD)/N$, and so is equivalent to R divided by the probability, as assessed before the

$$Pr(E|H) = \frac{c \times Pr(E|H_1) + Pr(E|H_2)}{c + 1}, \tag{An1}$$

where $c = Pr(H_1)/Pr(H_2)$. In the case being discussed here, $c = mD/N$; substituting for the general terms and multiplying numerator and denominator by N therefore yields Expression (A2).

Equation (An1) can be derived by recognizing that $Pr(H) \times Pr(E|H) = Pr(E \text{ and } H)$, that $Pr(H) = Pr(H_1) + Pr(H_2)$, and that $Pr(E \text{ and } H) = Pr(E \text{ and } H_1) + Pr(E \text{ and } H_2)$. Thus, we have

$$Pr(E|H) = \frac{Pr(E \text{ and } H_1) + Pr(E \text{ and } H_2)}{Pr(H_1) + Pr(H_2)}. \tag{An2}$$

The numerator of the fraction on the right side of Equation (An2) may be rewritten as $[Pr(H_1) \times Pr(E|H_1)] + [Pr(H_2) \times Pr(E|H_2)]$. Making this substitution, dividing numerator and denominator by $Pr(H_2)$, and substituting c for $Pr(H_1)/Pr(H_2)$, yields Equation (An1).

database search, that the source of the crime sample is not in the database given that Matcher is not the source.¹²⁷ This latter factor reflects the evidentiary value of the fact that the database search has eliminated as the possible source of the crime sample everyone represented in the database but Matcher.

C. *Posterior Probabilities*

Assume that, before the DNA evidence is received, every person represented in the database appears equally likely as each other to be the source of the crime sample, and every member of the suspect population not represented in the database appears equally likely as each other to be the source. Then the prior odds that Matcher is the source are

$$m/(mD + N). \quad (\text{A7})$$

The posterior odds equal the prior odds times the likelihood ratio, or Expression (A6) times Expression (A7), which simplifies to

$$Rm/N, \quad (\text{A8})$$

and to

$$R/N \quad (\text{A9})$$

on the assumption that $m = 1$, so that every member of the suspect population is as likely as every other, absent the DNA evidence, to be the source of the crime sample.

Now compare this last variation of the trawl case to a confirmation case, in which the prior odds that Matcher is the source of the crime sample are deemed to be

$$x/(D + N). \quad (\text{A10})$$

That is, Matcher is perceived to be x times as likely as the "average" member of the suspect population to be the source of the crime sample. Because the likelihood ratio in the confirmation case is simply R , the posterior odds are

$$xR/(D + N). \quad (\text{A11})$$

127. For a more general derivation, see Balding & Donnelly, *Evaluating DNA Profile Evidence*, *supra* note 13, at 607.

Expression (A11) is greater than Expression (A9) whenever $x/(D + N) > 1/N$,¹²⁸ or

$$x > 1 + (D/N). \quad (\text{A12})$$

If D is small in relation to N , then x has to be barely greater than 1 for Expression (A12) to hold. Even if D is as large as N , x has to be only greater than 2 — meaning that, absent the DNA evidence, Matcher has to appear only twice as likely as the average member of the population to be the source — for the posterior probability to be greater in the confirmation case than in the trawl case.

In speaking of the posterior odds in this analysis, we mean the odds assessed right after receipt of the DNA evidence. It is possible in the trawl case, of course, that the identification of Matcher will lead to other evidence that may raise those odds further.

D. *The NRC Analysis*

The analyses above have taken into account the fact that the profile matching the crime sample is that of a particular person, the one who thereby becomes of interest if he was not already. Now, for comparison, we will summarize, in terms similar to those used above, the likelihood ratio analysis of the trawl case offered by NRC II. Mistakenly, in our view, this analysis takes as the proposition at issue the generalized one that the source of the crime sample is someone represented in the database. It then assesses the likelihood ratio with respect to this proposition of the limited information that within the database there was exactly one match of the crime sample.

The numerator of this likelihood ratio is the probability that, if the source were someone represented in the database, exactly one profile in the database would match that of the crime sample. The source would certainly match. The probability that none of the other D samples would match is the same as (A1), $(1 - p)^D$.

The denominator of this likelihood ratio is the probability that, if the source were someone not represented in the database, exactly

128. In $x/(D+N)$, which is multiplied by R to form the posterior odds in the confirmation case, the numerator of x reflects the elevated prior odds, while the denominator of $(D+N)$ reflects the fact that the DNA evidence has not eliminated any other suspects. By contrast, in $1/N$, which is multiplied by R to form the posterior odds in the variant of the trawl case being considered here, the numerator of 1 reflects prior odds that did not distinguish Matcher from the rest of the population, while the denominator of N reflects the fact that the DNA evidence has eliminated D alternative suspects.

one profile in the database would match that of the crime sample. For any given sample, the probability that *that* sample, and no other samples in the database, would match the crime sample is $p(1 - p)^D$, the same as (A3). Because there are $D + 1$ samples in the database, the aggregate probability that one, and only one, profile in the database will match that of the crime sample, given the hypothesis that the source is actually somebody not represented in the sample, is therefore

$$(D + 1) \times p(1 - p)^D. \quad (\text{A13})$$

The likelihood ratio is therefore

$$\frac{1}{p \times (D + 1)}, \quad (\text{A14})$$

or

$$\frac{R}{D + 1}, \quad (\text{A15})$$

which is Expression (1) in the text.