Juror Understanding of Random Match Probabilities[1]

Dale A. Nance
Case Western Reserve University School of Law

The Science of DNA Profiling: An Expert Forum
Dayton, Ohio: August, 2007

Many of the debates that rage concerning expert testimony in general, and DNA evidence in particular, proceed from beliefs about how jurors react to such evidence. Some lawyers and commentators have great confidence in juries; others express great skepticism. In fact, the law of evidence is pervaded by numerous, often conflicting assumptions about jury competence. Nowhere is this ambivalence more conspicuous than when scientific evidence is quantified. On the one hand, it is generally accepted that jurors are capable of handling evidence that can be so quantified. On the other hand, there is considerable skepticism about jurors' ability to handle the quantification itself. It is useful, therefore, to examine what we know about jury behavior in regard to quantifiable scientific evidence.

In terms of serious social science, it will probably come as no surprise to you to learn that what we don't know about jury reactions to scientific evidence is probably much greater than what we do know. Nevertheless, we do know a good deal. I will limit my remarks here to a special but important topic, specifically, what we know about how jurors react to trace evidence (especially DNA evidence) reporting a match between the perpetrator and the accused accompanied by a quantifiable "random match" or "coincidental match" probability ("RMP"). This is just the probability that an innocent person tested for this attribute would share it with the perpetrator by coincidence. This probability is critically important to understanding the significance of a correctly conducted test. The easiest way to see this is to ask yourself this question: If you believe that the perpetrator and the accused both have a certain characteristic, X, does that tell you anything? If X is "red blood cells," which everybody has, then evidence that the perpetrator and the accused both have X is completely unhelpful, indeed irrelevant. If, on the other hand, X is some characteristic that only one person in the world could have, then that tells you a great deal. The random match probability is just the portion of the relevant population of potential suspects who share the characteristic in question. The smaller that proportion, the more probative is the fact of a match between the perpetrator and the accused.

Thus, I assume that such an RMP is meaningfully calculated with an acceptable degree of reliability for the case in question, although this can be highly controversial in the context of some types of forensic science, especially in regard to determining a "relevant population." Certainly, the most important modern example in which meaningful quantification is possible is DNA evidence, but other trace evidence analyses, such as typing under the ABO blood group system, present fundamentally the same issues, because they admit of quantified random match probabilities. I will refer throughout to DNA evidence, but I will discuss the results of empirical studies that considered other forensic science techniques that permit the calculation of a random match probability. I will limit my discussion to so-called "single stain" cases, in which there is only trace evidence from a single perpetrator. All the empirical studies of jurors (of which I am aware) have been conducted in this context.

The empirical studies I will be discussing are true "experiments," and I should take a moment to explain what I mean by that. I will *not*, for example, be talking about surveys of, or interviews with, jurors in real cases. Such surveys are notoriously unreliable, in that jurors may do one thing but say something quite different after the fact, whether intentionally or otherwise. Instead, I am referring to controlled experimentation in which a group of mock jurors is presented with a hypothetical case, with subgroups exposed to specific variations in the evidence presented. This is the epistemically preferred form of scientific inquiry, but it comes with costs. Because hypothetical cases are distanced from reality, there is the risk that mock jurors will not respond as they would in a real case, with persons' lives on the line. On the other hand, the advantages of such experiments are enormous, because it allows the experimenter to control, by randomization, for potential causal factors other than the variables of interest, thus dramatically reducing the chances of spurious judgments of causality.

I organize my comments around eight hypotheses that have been generated about how jurors will misinterpret or misuse the RMP. It is revealing that many researchers have focused on the risk of cognitive error by jurors. This reflects the prevailing suspicion of the competency of jurors, especially when it comes to scientific evidence. More specifically, it reflects the concern that jurors will be credulous in the face of expert testimony, naively accepting and thus overvaluing what the expert has to say, or what the jurors think the expert is saying. I will return to this theme at the end of my talk, after we have examined what we know about the frequency of the hypothesized cognitive errors. The first four of these hypothesized cognitive errors will, if they occur, generally favor the prosecution, which we will assume to be the proponent of the DNA match evidence. The last four will, if they occur, generally favor the defense.

Although the experiments that have been conducted have used various figures for the RMP, for presentational convenience I will mostly refer to a particular assumed value, namely, 1 in 40,000. That is much larger than is often true in DNA cases, but such figures do occur even in the DNA context.[2] Moreover, some of the explanations will be easier, for present purposes, than if I tried to work with an RMP of, say, 1 in a trillion. Conversely, the figure is generally smaller than the RMPs available for non-DNA evidence, though again there are exceptions. In any event, 1 in 40,000 is the figure that I used in an empirical study I conducted in Chicago between 2000 and 2002, so using that figure here will allow me to report in context some specific results from my study. At appropriate points, I will make comments about larger or smaller RMPs where the discussion requires it.

1. The Prosecutor's Fallacy

Suppose, then, an expert witness for the prosecution testifies about a DNA match, reporting an RMP of 1 in 40,000. Properly interpreted, this means only that there is a 1 in 40,000 chance that the defendant's DNA sample would match that of the crime scene

---

[2] Cf. R. v. Paul, 62 O.R. 3d 617 (Ontario Appeals, 2002) (involving random match probability of 1 in 58,000).

sample IF the defendant were, in fact, not the source of the crime scene sample. It is possible, however, that a juror will understand this to mean that there is a 1 in 40,000 chance that the defendant is not the source of the crime scene sample, or that there is a 1 in 40,000 chance that the defendant is innocent.

If the juror does understand this number that way, he has committed what Bill Thompson dubbed the "Prosecutor's Fallacy," and what is sometimes called "transposing the conditional" because it confuses the probability of innocence given the evidence with the probability of the evidence given innocence; the testimony is provided the latter, not the former. That the juror has erred in equating these probabilities is evident from what the juror's inference <u>fails</u> to take into account. The report of a match could be the result of false positive lab error, so the chance that the samples are not from the same source is considerably larger than 1 in 40,000. For example, one systematic review of the results of several lab proficiency tests conducted before 1995 reached an estimate of 1 error in every 800 tests. (Koehler, et al., 1995.) Further, the incriminating evidence might have been planted by the police, consideration of which reduces not the probability that they are from the same source but rather the probability that, if they are from the same source, the defendant is innocent nonetheless. These other explanations of a reported match may or may not be very significant possibilities in context, but the simple inference described above has not taken them into account at all, because the juror's inference as stated does not utilize any evidence in the case except the fact of a match and the random match probability.

So how likely is it that a juror will reason this way? By counting how often jurors equate the probability of innocence with the RMP (or the probability of guilt with (1−RMP)), one can obtain an approximate, if generous, measure of the frequency of this fallacy. The available experiments, including my own, have produced estimates ranging from well below of 1% of jurors to over 20% of jurors, prior to deliberations. The differences in these results appear to arise from several factors. These studies suggest that this particular error may be more common the larger the RMP (e.g., more common with an RMP of 1 in 25 than an RMP of 1 in 1,000,000), more common when the RMP is expressed as a conditional probability rather than as a frequency of occurrence (e.g., "the chance of coincidental match is 1 in 40,000" rather than "about 1 out of every 40,000 people share this profile"), and more common if arguments of counsel do not cause jurors to focus on the fallacy. (Thompson & Schumann, 1987; Goodman, 1992; Nance & Morris, 2002; Nance & Morris, 2005.)

How great must the incidence level be before we should be seriously concerned about a jury verdict being tainted? The answer depends on the effect of jury deliberations. There is no empirical work directly addressing this point, but some plausible reasoning may help. Presumably, fallacious reasoning would be at a distinct argumentative disadvantage in deliberation. In fact, one study presented mock jurors with an argument *expressly making* the fallacious inference, and found that so doing actually *reduced* the incidence of the fallacy among the jurors. (Thompson & Schumann, 1987, Experiment 2.) So the question becomes: How likely is it that a certain fallacy is so common in a particular jury that those not suffering from it will be unable to dissuade the others from their error? Based on the incidence rate, one can calculate probabilities that any given number of randomly drawn jurors will fall for the fallacious inference. Using such calculations, I have argued that an incidence rate as high as 5% poses no serious threat to the integrity of a post-deliberation verdict, an estimate that is probably very conservative – that is, very risk averse – given the probable effect of arguments that would focus on the fallacy. (Nance & Morris, 2002.) Such a low rate can probably be achieved by communicating RMPs as frequencies rather than probabilities, whether this is done by requiring this form in direct testimony or by eliciting it during cross-examination. Of course, this does all assume that some attention is given to the issue by counsel.

2. Neglect of Lab Error

In recent years, controversy has attended the question of the rational dominance of the chance of laboratory error over the chance of coincidental match, in contexts where the former is much larger than the latter. A random match probability for DNA evidence that is on the order of, say, one in a billion is considerably less impressive than it might seem if one takes into account a much larger false positive laboratory error rate of, say, one in a hundred, or even one in a thousand. In such a context, the combined chance of a report of a match due to *either* a coincidental match *or* a false positive lab error, assuming the accused is innocent, is essentially the same as the chance of lab error; the chance of coincidental match is completely swamped by the chance of lab error, and the rational trier of fact would, in this sense, ignore the random match probability and focus only on the chance of lab error. It has been argued that if jurors do not understand this, there is a risk that they will greatly overvalue the match evidence. (Koehler et al., 1995.) One hypothesized cause of such misunderstanding is that, unless the chance of lab error is quantified for the jury, using estimates derived from proficiency tests, the jury will treat the case as if there were no chance of lab error. (Schklar & Diamond, 1999.) Jurors might simply fail to attend to the chance of lab error, or they might mistakenly think that the RMP incorporates the lab error risk. If so, then they will attribute too much significance to a small RMP, and thus give too much weight to the evidence of a match.

Does this happen? Using an assumed figure for lab error rate (1 in 1,000) derived from various proficiency studies, and a much smaller RMP (1 in 40,000), I conducted an experiment that tested whether giving or not giving testimony reporting such a lab error rate had any effect on mock jurors' rated probabilities of guilt or their willingness to convict the defendant. Even with a very large sample of mock jurors, making a statistically powerful test, no effect was detected. (Nance & Morris, 2005.)[3] The probable explanation lies in the fact that, even when jurors were not given a quantified estimate of the lab error rate, they were aware of the possibility of such error and that it is not part of the random match probability. In fact, without any testimony quantifying the lab error rate, jurors' median estimate of the chance of lab error (1 in 1,000) was the same as the estimate derived from proficiency studies. It appears that jurors are not only aware of the chance of lab error, even if it is not quantified, but they have a surprisingly good sense of how likely it is.

---

[3] In my experiments, the prosecution's expert did acknowledged that there was a chance of a lab error even in those conditions in which no quantified estimate of such error was provided. It is possible, therefore, that jurors might neglect lab error in cases where the prosecution experts do not even acknowledge the *possibility* of error, provided the defense is unable or unwilling to impeach such unrealistic confidence.

Does this mean we shouldn't bother with testimony about lab error rates? I don't think so. First of all, if a particular lab has an unusually high or low error rate, communicating that rate to the jurors might have an effect, although I have not seen this tested experimentally. Second, and perhaps more important, presenting the evidence so that the jury knows the lab error rate has been taken into account in some fashion may also affect juror reactions. This has been tested, and I will discuss it later in my talk.

3. Averaging and Other Improper Combination Strategies

Another cognitive mistake with regard to the importance of lab error has been hypothesized. Failing to inform jurors how to combine the lab error risk with the random match risk, when both are quantified in testimony, might cause them erroneously to *average* the two risks instead of *adding* them, thus underestimating the combined risk of a false indication of guilt in the testimony. Based on this risk, one prominent researcher in this area, Jonathan Koehler, has argued that jurors should not be given separate estimates of the RMP and lab error rate, but should only be given an estimate of the combined risk, in which case jurors would never hear numbers like 1 in a billion. (Koehler et al., 1995; Koehler, 1997.)

But does such improper averaging occur to any significant extent? The experimental evidence on this question is somewhat conflicting. One experiment, conducted by Chicago researchers Shari Diamond and Jason Schklar, asked mock jurors *not* provided with combination instructions to provide an estimate of the combined risk. The researchers reported that a majority of the jurors used either averaging or some other form of improper combination strategy that would understate the combined risk. On the other hand, if these errors were to have an effect on the jurors' decision, one would expect higher rated probabilities of guilt and willingness to convict rates when only separate estimates of RMP and lab error rate are given (thus inviting the error) than when correct combination instructions are also given in testimony (discouraging the error). Yet, these researchers found that the jurors' verdict preferences were not significantly affected by their being given correct combination instructions. (Schklar & Diamond, 1999.) Of course, this non-finding might have been due to an inadequate sample size (they had relatively small subgroups) or the nature of the mock jurors employed. Like Koehler, Diamond and Schklar used readily available undergraduate students as their test subjects.

My later research attempted to shed light on this by looking at both the jurors' assessments of the probability of guilt and their verdict preferences. My study also involved much larger sample sizes, and they were drawn from a real jury pool rather than from a group of undergraduate students. I found that only in one special context did giving a combination instruction have a statistically significant effect on rated probabilities of guilt or verdict preferences, and in that context the result was the opposite of what the averaging hypothesis would predict: jurors given a combination instruction provided *higher* probability of guilt assessments and a *greater* proportion of such jurors were willing to convict. (Nance & Morris, 2005.) I will come back to this intriguing result later.

Thus, while it does appear that some people have trouble performing the combination without instruction, it is not clear that such difficulties have a discernible impact on jurors' verdicts. This could be because jurors make better *use* of the information than they are able to *demonstrate* by explicit computation. This is not an unusual phenomenon. It might also be because other risks – such as the risk of police mistakes or intentional planting of the incriminating evidence – are dominant in the jurors' assessments of the probative value of the DNA match. In three studies, mock jurors have provided their own untutored estimates of the chance of such causes of a match report; the mean estimate is about 1 in 50 (0.02) and the median estimates are between 1 in 500 (0.002) and 1 in 1,000 (0.001). Indeed, in these studies, jurors' estimates of the risk of such events are generally somewhat higher than the jurors' estimates of the risk of lab error. (Schklar & Diamond, 1999; Nance & Morris, 2002; Nance & Morris, 2005.)

4. The Vividness Hypothesis

A fourth reason has been suggested to warrant the inference that jurors overvalue DNA match evidence when a small RMP is presented. The idea is that when a very small number, like 1 in a billion, is presented to a juror, its *vividness*, or *memorability*, will cause the juror to focus only, or disproportionately, on that small number, ignoring more pallid, but rationally more significant, numbers like a lab error rate on the order of 1 in 100 or 1 in 1,000.

The main evidence that has been offered in support of this hypothesis is that willingness to convict rates increase substantially when a small RMP is introduced, as compared to experimental conditions in which no RMP is provided in the testimony. What's more, this holds true even if the condition in which no RMP is provided is one in which mock jurors are given a risk rate that they are told combines the risk of lab error and the risk of coincidental match. Again, Jonathan Koehler conducted these interesting experiments. (Koehler et al., 1995.) These facts are certainly consistent with the hypothesis of a vividness effect, but there are a number of alternative explanations of the data.

The most obvious is that jurors are perplexed by the absence of an explicit estimate of the RMP; because they may have come to expect some such estimate, based on their familiarity with such cases in the news, they might even be suspicious of the prosecution's failure to present the RMP. Similarly, if told that a given figure, like 1 in 1,000, combines the risk of coincidental match and the risk of lab error, they might be concerned that the two figures are not broken out for them so that they could assess them directly. Or in some other way, jurors might be perplexed by the absence of the RMP or at least data from which the frequency of occurrence of the DNA profile in the population might be inferred. They might, in fact, be concerned that something is being hidden from them, as indeed it is. If so, they might discount their assessed probability of guilt and be *less* willing to convict when the RMP is not quantified. In other words, it may be that evidential incompleteness (by omitting the RMP) causes *under*valuation rather than that vividness (by including the RMP) causes *over*valuation. Koehler did not attempt to develop normative measures of where the jurors should come out, so his conclusion that vividness was error *increasing* was not supported by his results. A vividness effect, if that is what was occurring, might instead have been error *reducing*.

Two other studies (one by Shari Diamond one of mine) shed light on this issue. Both found that, assuming a vividly small RMP is presented, a *higher* willingness to convict is manifested when a more pallid (less impressive) lab error estimate is also presented than when it is omitted. (Schklar & Diamond, 1999; Nance & Morris, 2005.) These results tend to undermine the vividness hypothesis by showing that jurors still attend to the less impressive number. Moreover, they tend to support a competing "incompleteness" hypothesis, namely, that jurors are more willing to convict when an otherwise vague risk of error is meaningfully quantified.[4] I will come back to this point.

5. The Defense Attorney's Fallacy

We turn now to the hypothesized cognitive errors that, if present, would tend to favor the defense. What has been called "the defense attorney's fallacy" arises when a juror infers that the match testimony is *irrelevant* because there are too many people in the population who share the DNA profile of the accused and the perpetrator. Thus, with a 1 in 40,000 RMP and a just plausible suspect population of, say, 12,000,000, a juror might infer that about 300 other people could be the perpetrator and *therefore* that the match evidence is irrelevant. The mistake, of course, is in thinking that a single item of evidence is irrelevant unless it, by itself, narrows the suspect population down to one person or at least a very small number of persons. By such a standard, evidence about most visible characteristics of a perpetrator, such as gender, skin color, or hair color, that match the defendant would be irrelevant, indeed inadmissible.

Experimental evidence suggests that this cognitive error is comparatively common, with estimates ranging from 6.5% to 66% of jurors affected, at least without the benefit of deliberations. (Thompson & Schumann, 1986; Nance & Morris, 2002.) The higher range frequency estimates for mock jurors arose in the presence of unchallenged arguments expressly encouraging the fallacy. Even without such arguments, however, the estimates range up to 12%. Thus, the experimental evidence suggests that the occurrence of this fallacy may be a serious problem affecting the accuracy of jury decision making, and that it can be exacerbated by arguments of counsel pushing jurors in that direction.

---

[4] Moreover, outside the context of forensic science, the general psychology literature reveals an accumulation of evidence against the proposition that vividness affects decision. (Taylor & Thompson, 1982; Collins, Taylor, Wood, & Thompson, 1988.)

Indeed, until fairly recently, even some courts fell into this error in rulings on the admissibility of forensics, at least when the RMP was large.[5] As this would suggest, the fallacy is less likely to occur if the plausible suspect population is small enough in relation to the RMP that the juror will not see the matching characteristic as one shared by a large number of potential perpetrators. Thus, an RMP of 1 in a million is not likely to give rise to the defense attorney's fallacy if a typically "local" crime, like rape, occurred in a town with a population of less than one million, and an RMP of 1 in a trillion is not likely to give rise to the defense attorney's fallacy in any case. In the context of more moderate RMPs and large potential suspect populations, it is important for prosecutors to be prepared to counter this tendency with argument or testimony.

6. The Defense Attorney's (Extreme) Fallacy

A more extreme form of the Defense Attorney's fallacy occurs when an incriminating match report is considered relevant but *exculpatory* because, it is reasoned, if about 300 people in the suspect population can be expected to share the DNA in the profile, then the probability is only 1 in 300 that the defendant is guilty. Like the Prosecutor's Fallacy, this is fallacious because it ignores all other evidence in the case. It is like saying that, if the prosecution offers, *among other evidence*, testimony that the perpetrator, like the defendant, is male, and there are 100,000 males in the community, then the probability of guilt must be 1 in 100,000; it's true only if all other evidence is ignored.

As far as I am aware, no direct empirical tests have yet measured the frequency of this form of the fallacy. It is suggested by Koehler's finding that roughly 40% of mock jurors in a DNA case with moderate a RMP (1 in 1,000) set in a large urban area (Houston) rated the source probability or guilt probability at extremely small levels, less than a 1% chance. (Koehler, 2001a, Experiment 3.) However, this result probably does not indicate the occurrence of the extreme fallacy for one important reason: there was

---

[5] See People v. Mountain, 486 N.E.2d 802, 805-06 (N.Y. 1985) (overruling precedent that would exclude evidence that both perpetrator and defendant had type "A" blood, with a random match probability of 0.4 because 40% of the population had type "A" blood).

virtually no evidence offered against the accused other than the DNA match. In that context, the reasoning described is not actually fallacious. In fact, Koehler's point about this result was not that it indicates a fallacy, but rather that these defense favorable responses from jurors occur when the evidence is presented so as to facilitate jurors' ability to imagine coincidental matches (as when the testimony states that "1 in 1,000 people in Houston would also match"), but not when the evidence focuses only on the defendant (as when the testimony states that "the probability that the suspect would match if he were not the source is 0.1%.")

In any event, in the context of DNA evidence with random match probabilities that are very small relative to the inverse of the size of the suspect population, this would appear to be a minor risk. For example, Koehler found that with a smaller RMP (1 in 1,000,000 or smaller) in the same urban area (Houston), the proportion of jurors who estimated the probability of guilt at less than 1% dropped to about 3%, and their was relatively little difference in responses given to the two presentation formats he used. But in those cases involving relatively large RMPs, prosecutors may need to try to guard against this fallacy by appropriate testimony or argument.

7. The Inversion Fallacy

Recall that the Prosecutor's Fallacy improperly equates the RMP with the probability of innocence. Another mistake, one that typically favors the defense, is to equate the RMP with the probability of guilt. This bizarre mistake causes the rated probability of guilt to be inversely related to the rational probative value of the match because the smaller the RMP the greater the rational probative value of a match, but the *smaller* the probability of guilt attributed by someone falling into this error. Hence the name.

You might think that no one would make this mistake. But you would be surprised. In my research, this error was found to be surprisingly prevalent in the context of a relatively high RMP of 1 in 25. In that context, about 5% of jurors gave responses suggesting the inversion fallacy. (Nance & Morris, 2002.) With a much smaller RMP of 1 in 40,000, however, only one juror among 1,205 could be clearly identified as probably reflecting this mistake. (Nance & Morris, 2005.) To be sure, this might understate the incidence of this mistake, because someone making it might simply report the probability of guilt as 0%, rather than bothering to say 1 in 40,000. In fact, 35 jurors (3% of 1,205) gave probability of guilt ratings of 0%.[6] This figure, however, likely overstates the incidence of this error, because more than 7% of the jurors in that study who did not receive any quantified RMP, from which this error could occur, also rated the probability of guilt as 0%. Thus, some jurors attribute no weight to such evidence but do so without making the inversion fallacy. In any event, because this error is so egregious, it must be particularly amenable to correction by deliberation with jurors not so affected.

8. Misaggregation

The last cognitive error is the most difficult to describe and to assess, but it is likely the most important of the bunch. Unlike the other defense-favoring cognitive errors, this one does NOT generally involve the juror ignoring the other evidence in the case. Rather, it involves the juror using an inaccurate method of combining the effect of that other evidence with the DNA evidence. This requires some explanation.

Without regard to the DNA match evidence, jurors have some sense about the likelihood of the defendant's guilt. The communication of an RMP is designed to provide jurors with information necessary to revise their assessment of the probability or odds of guilt in light of the evidence of a DNA match. But it has been known, or at least believed, for some time that ordinary people tend to be more conservative, when revising assessments of probability in light of new quantitative information, than the rules of probability prescribe. (Edwards, 1968.) If one takes these rules of probability as normative, as suggesting measures of how jurors *ought* to come out in a case, the cumulative import of several of the experiments with forensic match evidence and quantified RMPs is that people have difficulty giving statistical evidence as much weight as it deserves, a problem called "misaggregation." (Faigman & Baglioni, 1988;

---

[6] In another study, the author suggested that, in the context of an RMP of 1 in a billion, the incidence of this mistake might be estimated by the percentage of respondents giving any probability of guilt less than 1%, or even less than 10%; the corresponding incidence rates were 4% and 8%, respectively. (Koehler, 2001b.)

Goodman, 1992; Smith, et al., 1996; Taroni & Aitken, 1998; Schklar & Diamond, 1999; Nance & Morris, 2002; Nance & Morris, 2005.)

The two most important observations about this otherwise consistent result concern the *degree* to which jurors undervalue the evidence. First, the apparent magnitude of this effect has declined as experimenters have become more sophisticated in their determinations of the norm against which the jurors' assessments are measured. Calculating these norms – using what is known as Bayes' Rule – is complicated business, and early attempts did not take into account the various ways, other than a coincidental match, that a witness might report a forensics match when a defendant is in fact innocent. (Thompson & Schumann, 1987; Faigman & Baglioni, 1988; Goodman, 1992; Smith, et al., 1996; Taroni & Aitken, 1998). In other words, the jurors probably were, quite properly, providing estimates of the probability of guilt (or the probability that the two samples were from the same source) that took into account certain risks of convicting the innocent (such as lab error or police fabrication) that the experimenters had not taken into account in calculating a normative measure. When such other sources of a falsely incriminating match report are taken into account, there does still appear to be undervaluation, but to a much lesser degree. (Schklar & Diamond, 1999; Nance & Morris, 2002; Nance & Morris, 2005.)

The second, and surely more controversial, issue concerns the extent to which variation in the way that the DNA evidence and the RMP are presented and explained to jurors affects the degree of undervaluation and how we can use knowledge of this phenomenon to reduce the extent of under-valuation. Two main strategies find some basis in the literature. First, as already noted, when the random match probability is given in testimony as a conditional probability focused on the defendant (e.g., "The probability that defendant, if innocent, would match the perpetrator's DNA profile is 1 in 40,000") rather than as a frequency focused on the population (e.g., "On average, 1 in every 40,000 people have the same DNA profile as the perpetrator and the defendant") jurors tend to rate the probability of guilt higher and to be more willing to convict. (Koehler, 2001a.) The study reporting this effect did not, however, attempt to calculate Bayesian norms for the hypothesized case, against which to compare the jurors' responses, so it cannot be determined which of these presentation formats was more accurate, even in Bayesian terms. Moreover, as noted above, use of the conditional probability format may encourage the Prosecutor's Fallacy or conceal the chance of coincidental match, so that any improvements in accuracy obtained by using that format might be purchased at the cost of inducing fallacy. That is, jurors might be "tricked" into more accurate verdicts, surely a dubious kind of "improvement."

A second strategy for reducing the misaggregation effect is to provide jurors with instruction in the mathematics of Bayes' Rule, for example, by providing them with a chart that associates various prior probabilities (the probability of the defendant being the source or being guilty without regard to the DNA match evidence) with various posterior probabilities (probabilities taking the DNA match into account) that Bayes' Rule prescribes. Here is an example, for a random match probability of 1 in 40,000 (ignoring sources of error):

Prior Probability	→	Posterior Probability
0 %	→	0 %
1 in 1 million	→	3.8463 %
1 in 500,000	→	7.4074 %
1 in 100,000	→	28.5716 %
1 in 10,000	→	80.0016 %
1 in 1,000	→	97.5634 %
1 in 100 (1 %)	→	99.7531 %
1 in 10 (10 %)	→	99.9775 %
20 %	→	99.9900 %
30 %	→	99.9942 %
40 %	→	99.9963 %
50 %	→	99. 9975 %
60%	→	99.9983 %
70 %	→	99.9989 %
80%	→	99.9994 %
90 %	→	99.9997 %
100 %	→	100 %

This idea, suggested in an article in the Harvard Law Review nearly 40 years ago (Finkelstein & Fairley, 1970), triggered a major controversy that has continued ever since.

Does this work? Early studies attempting to assess this suggestion, using mostly students as test subjects, found little effect of such instruction. (Faigman & Baglioni, 1988; Smith, et al., 1996.) However, my recent experiments, using much larger sample sizes and citizens drawn from the jury pool of Kane County, Illinois, found clearly to the contrary: Bayesian instruction increased the jurors' assessed probability of guilt, in the direction of Bayesian norms, and their willingness to convict. This was true with both relatively large RMPs (like 1 in 25) and relatively small RMPs (like 1 in 40,000), and it was true despite the fact that the RMP was commuicated as a frequency, in a manner that would facilitate jurors imaging coincidental matches. (Nance & Morris, 2002; Nance & Morris, 2005.)

One of the most interesting aspects of my results is something I have alluded to already: when explicit estimates of lab error were incorporated into the Bayesian calculations shown to jurors, their assessed probability of guilt increased further, approaching the Bayesian norm, and their willingness to convict was the highest of all experimental conditions examined. (Nance & Morris, 2005.) In some respects, this is a surprising result. Instruction that incorporates lab error communicates *smaller* posterior probabilities, yet the jurors' assessed posterior probabilities were *larger*. This tends to confirm what I earlier described as the "incompleteness" hypothesis: that jurors discount even powerful scientific evidence when something is missing that prevents them from having confidence that they understand the import of the evidence that is presented. Bayesian instruction that doesn't incorporate the lab error leaves them wondering how to account for lab error, and in default of instruction about this, they discount the probability of guilt.

So what is the "big picture" here? The experimental evidence reveals mock jurors, both students and citizens from jury pools, who are not credulously accepting whatever the expert seems to be saying. Instead, it shows jurors who are cautious and skeptical of expert testimony. To be sure, one cognitive error favoring the prosecution, the Prosecutor's Fallacy, can pose a serious risk, but it can be ameliorated by testimony giving the RMP in terms of frequencies (such as "1 in every 40,000 people on average share this DNA profile") instead of conditional probabilities (such as "the probability that the accused would match this profile by coincidence is 1 in 40,000") or by argument from defense counsel that reframes the RMP in terms of frequency of occurrence in the population.

Less tractable are cognitive errors favoring the defense. The Moderate and Extreme versions of the Defense Attoney's Fallacy may be quite common, especially when defense counsel encourage it. We do not know as yet to what extent legitimate counter-arguments by the prosecution can ameliorate these errors, but with small RMPs, especially those on the order of 1 in a billion, these fallacies should be rare. So, from what we know so far, probably the most important and persistent cognitive error affecting even cases with a very small RMP is misaggregation, the failure to give statistical evidence the weight it deserves in the case. This suggests that, rather than focusing our energies on excluding parts of the evidence thought to cause jurors to go astray, like infinitesimal RMPs, we should instead be focused on finding better ways to explain the significance of the RMP, however large or small it may be.

This conclusion is consonant with the implications of empirical studies of jurors in other settings not involving scientific evidence. Several lines of converging research are confirming the general principle that educating jurors about problems is generally a better strategy than trying to blindfold them.

References

Collins, R., Taylor, S., Wood, J., & Thompson, S. (1988). The Vividness Effect: Elusive or Illusory? *Journal of Experimental Social Psychology, 24*: 1-18.

Edwards, W. (1968). Conservatism in Human Information Processing. In B. Kleinmutz (Ed.), *Formal Representation of Human Judgment* (pp. 17-52). New York: Wiley.

Faigman, D., & Baglioni, A. (1988). Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence. *Law and Human Behavior*, 12:1-17.

Finkelstein, M. & Fairley, W.B. (1970). A Bayesian Approach to Identification Evidence. *Harvard Law Review, 83*: 489-517.

Koehler, J., Chia, A., & Lindsey, S. (1995). The Random Match Probability in DNA Evidence: Irrelevant and Prejudicial? *Jurimetrics Journal, 35*: 201-219.

Koehler, J. (1997). Why DNA Likelihood Ratios Should Account for Error (Even When A National Research Council Report Says They Should Not). *Jurimetrics Journal*, 37: 425-437.

Koehler, J. (2001a). When Are People Persuaded by DNA Match Statistics? *Law and Human Behavior, 25*: 493-511.

Koehler, J. (2001b). The Psychology of Numbers in the Courtroom: How to Make DNA Match Statistics Seem Impressive or Insufficient, *Southern California Law Review, 74*: 1275-1305.

Nance, D. & Morris, S. (2002). An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Large and Quantifiable Random Match Probability. *Jurimetrics Journal, 42*: 403-445.

Nance, D. & Morris, S. (2005). Jury Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random Match Probability. *Journal of Legal Studies*, 34: 395-444.

Schklar, J. & Diamond, S. (1999). Juror Reactions to DNA Evidence: Errors and Expectancies. *Law and Human Behavior*, 23: 159-184.

Smith, B., Penrod, S., Otto, A., & Park, R. (1996). Jurors' Use of Probabilistic Evidence. *Law and Human Behavior*, 20: 49-82.

Taroni, F., & Aitken, C. (1998). Probabilistic Reasoning in the Law, Part 1: Assessment of Probabilities and Explanation of the Value of DNA Evidence. *Science & Justice*, 38(3): 165-177.

Taylor, S. & Thompson, S. (1982). Stalking the Elusive Vividness Effect. *Psychological Review*, 89: 155-181.

Thompson, W. & Schumann, E. (1987). Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy. *Law and Human Behavior*, 11: 167-187.