

To appear in: *Southern California Law Review* (2001).

**THE PSYCHOLOGY OF NUMBERS IN THE COURTROOM:
HOW TO MAKE DNA MATCH STATISTICS SEEM
IMPRESSIVE OR INSUFFICIENT**

Jonathan J. Koehler¹

February 26, 2001

¹ Associate Professor, McCombs School of Business and School of Law, The University of Texas at Austin. This research was supported by National Science Foundation Grant SBR 9819305 "Understanding jurors' use of highly diagnostic statistical evidence." The author may be contacted at CBA 5.202, The University of Texas at Austin, Austin, TX 78712-1175, 512-471-7856, koehler@mail.utexas.edu. I thank Russell Korobkin, Thomas Lyon, and John Monahan for comments on a previous draft. I also thank Hal Arkes, Daniel Gilbert, Laura Macchi, Molly Mercer, Paul Slovic, Pat West, members of the Behavioral Economics seminar at Harvard University, and students in the STA380 seminar at The University of Texas, for helpful theoretical discussions along the way. Finally, I thank Arienne Brint, Andrew Gershoff, Nir Liberboim, Caryn Thesing, and Cherilyn Young for superb research assistance.

By now, everyone knows that forensic DNA analysis represents a stunning theoretical advance for the criminal justice system.² The value of DNA analysis lies in its theoretical power to exclude large proportions of the population as potential contributors of genetic samples (e.g., blood, semen, hair). Thus, when a suspect's DNA matches a DNA sample that is recovered from the scene of a violent crime, a prosecutor may suggest that the suspect is a likely source of that sample (and was therefore present at the crime scene) because the suspect is among the few who were not excluded as potential contributors. In most cases, the probative value of a DNA match is conveyed to triers of fact via frequency statistics that describe how common the DNA profiles are in a given population.³ A DNA match statistic⁴ of, say, 1 in 1,000,000 means that

² *New York v. Wesley*, 140 Misc. 2d 306, 533 N.Y.S.2d 643, 644 (Sup. Ct. 1988) (DNA analysis represents "the single greatest advance in the 'search for truth,' and the goal of convicting the guilty and acquitting the innocent, since the advent of cross-examination").

³ Identifying the appropriate population for comparison is a difficult problem that is often glossed over in practice. That population should include all people who may have contributed the genetic sample and exclude those who could not have contributed the sample. This has been referred to as the "potential source population," Jonathan J. Koehler, *DNA Matches and Statistics: Important Questions, Surprising Answers*, 76 JUDICATURE 222, 227 (1993); Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence*, 34 JURIM. J. 21, 26 (1993). This population will usually (but not always) be similar to the group of people who could have committed the crime (i.e., the "suspect population"), Richard Lempert, *The Suspect Population and DNA Identification*, 34 JURIM. J. 1, 3-4 fn. 8 (1993) (arguing that the distinction between the potential source population and the suspect population can usually be ignored because "in most cases the group of plausible DNA sources and the group of plausible suspects will be identical"). Although the suspect population will sometimes be too narrow (i.e., it will may exclude non-suspects who nonetheless may be the source of recovered genetic material), this reference group is widely accepted as a reasonable population from which to identify the frequency of DNA profiles, National Research Council, *DNA TECHNOLOGY IN FORENSIC SCIENCE* (1992) at S-4 ("What is the probability that such a match would have occurred between the suspect and a person drawn at random from the same population as the suspect? Answering that question requires calculation of the frequency in the population of each of the gene variants (alleles) that have been found, and the calculation requires a databank where one can find the frequency of each allele in the population") [hereinafter, NRC I]; Richard Lempert, *Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to the Reverend Bayes*, 13 CARDOZO L. REV. 303, 310 (1991); Richard C. Lewontin, *Which Population?*, 52 AM J. HUM. GENETICS 205 (1993) (Letters to the Editor) ("[T]he ethnicity of the defendant is not the directly relevant question, but rather the ethnic composition of the pool of possible alternative suspects"); Peter Donnelly and Richard Friedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 MICH L. REV. 931, 947 (1999) ("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").

⁴ Here I note that statistics are used to characterize inclusions and not exclusions. When no match is found between

approximately one person out of every one million in a population will match that DNA profile. An equivalent way to say this is that the chance that a randomly selected person will match this DNA profile is 1 in one million. To the extent that the DNA match statistic is low (e.g., 1 in many thousand, million, or billion), it is unlikely that the match between a suspect and a recovered crime scene sample is purely coincidental.

The early forensic DNA years were marked by controversy over the proper computation of DNA match statistics.⁵ Although disagreement has abated,⁶ little is known about how jurors think about and use DNA match statistics.⁷ It is widely assumed that DNA statistics are persuasive. That is, people assume that, after hearing that (a) a suspect matches traces of DNA evidence from a violent crime scene, and (b) the chance that a randomly selected person from the population would match is one in a million or billion, jurors will be convinced that the matching

the suspect and a genetic sample, the suspect is excluded as a possible donor of the sample and statistics are not provided.

⁵ For a review, see William C. Thompson, *Evaluating the Admissibility of New Genetic Identification Tests: Lessons From the "DNA War,"* 84 J. CRIM. LAW & CRIMINOL. 22, 61-89 (1993). The FBI has now disposed of statistics altogether in cases where the match statistics are extremely small. In these cases the FBI simply assumes that the DNA profile is unique and that no one other than the matching suspect could possibly be the donor. See Roberto Suro, *DNA Now Used To Make Specific Identification; FBI Calls Lab Match 'Major Breakthrough'*, WASH. POST, Nov. 13, 1997, at A4. Some DNA laboratories have implemented or are considering implanting a similar policy (personal communication with Irma Rios (Director of the Department of Public Safety DNA Laboratory, Austin, TX, January 24, 2000). For a thoughtful examination of the flaws inherent in the assumption of uniqueness in forensic science see Michael J. Saks, *Merlin and Solomon: Lessons from the Law's Formative Encounters with Forensic Identification Science*, 49 HASTINGS L. J. 1069, 1082(1998).

⁶ Eric S. Lander and Bruce Budowle, *DNA Fingerprinting Dispute Laid to Rest*, 371 NATURE 735, 735 (1994) ("We recently discussed the current state of DNA typing, and could identify no remaining problem that should prevent full use of DNA evidence in any court. What controversy existed seems to have been fully solved by the NRC report, the TWGDAM guidelines and the extensive scientific literature"). For an alternative perspective, see Laurence D. Mueller, *The Use of DNA Typing in Forensic Science*, 3 ACCOUNTABILITY IN RESEARCH 55 (1993).

⁷ National Research Council, COMMITTEE ON DNA FORENSIC SCIENCE: AN UPDATE, THE EVALUATION OF FORENSIC DNA EVIDENCE (1996) at 203 [herein after NRC II]; "we know very little about how laypersons respond to DNA evidence and how to minimize the risk that they will give the DNA evidence inappropriate weight"). See also Margaret A. Berger, *Laboratory Error Seen Through the Lens of Science and Policy*, 30 U.C. DAVIS L. REV. 1081, 1094 ("Virtually no empirical studies have as yet been conducted on juror reactions to extremely low probabilities").

suspect is indeed the source of the DNA.⁸ However, my own research with mock jurors indicates that the story is more complicated. Specifically, the way in which DNA match statistics are framed and presented to legal fact finders may affect how they think about and use the DNA evidence.

The possibility that the form in which jurors hear about the incidence of a matching DNA profile might affect how jurors perceive the probative value of DNA evidence captured the attention of a 1996 National Research Council [NRC II].⁹ After noting that “scientifically valid testimony about matching DNA can take many forms,” NRC II called for behavioral research on “the ability of jurors to understand the significance of a match as a function of the method of presentation.”¹⁰ One member of NRC I made a similar plea years ago.¹¹

Consider the following four formulations of a one in one thousand DNA match statistic that might appear in a criminal case:

1. The probability that the suspect would match the blood drops if he were not the source is 0.1%;
2. The frequency with which the suspect would match the blood drops if he were not the source is 1 in 1,000;
3. 0.1% of the people in Houston who are not the source would also match the blood

⁸ Harriet Chiang, *Court Evidence Controversy, Jury Still Out on Effectiveness of DNA Analysis*, SF CHRON. p. A1 (July 5, 1994) quoting San Francisco public defender Peter Keane (“when you hear one in a billion ... it's almost like the dead person sitting up, pointing and saying, 'He did it!'”).

⁹ This second National Research Council panel on DNA evidence was convened largely to address unresolved statistical issues.

¹⁰ NRC II, at ES-7.

¹¹ Richard Lempert, 1991, *supra* at 307 (“Proper weighing of DNA evidence by the jury requires a proper understanding of what the evidence imports. To achieve this a court must carefully monitor the way in which statistics are presented and interpreted by both expert witnesses and counsel”).

drops;

4. 1 in 1,000 people in Houston who are not the source would also match the blood drops.

These four formulations are legitimate, mathematically comparable ways to describe a one in one thousand DNA match statistic. However, they are psychologically different and their effect on jurors varies significantly. The bottom line appears to be this: When the statistic is framed in the language of probability (e.g., 0.1%) in a way that highlights a particular suspect's chance of matching by coincidence, it tends to be persuasive. But when the statistic is framed in the language of frequencies (e.g., 1 in 1,000) in a way that highlights the chance that others will match by coincidence, it is much less persuasive. Similarly, DNA match statistics that target an individual suspect are more persuasive than the equivalent statistic that those that target a broader population. Regarding the four formulations above, the first formulation provides the most convincing evidence, and the fourth formulation is least convincing. Indeed, as documented in the studies below, whereas a majority of jurors who are provided with the first formulation are at least 99% certain that the suspect is the source of the evidence sample, nearly as many who are provided with the fourth formulation are equally convinced that the suspect is *not* the source! In short, DNA match statistics that seem impressive when presented one way seem insufficient when presented another way.

This article describes a series of studies that I conducted with hundreds of jury-eligible¹² subjects in Austin, Texas. The studies examine factors related to the presentation of DNA statistics that affect whether a juror will be impressed with and persuaded by the DNA evidence.

The data are described in the context of a novel psychological theory for predicting when legal decision makers will and will not be impressed and persuaded by statistical evidence. Section I lays out the theory of interest, "exemplar cueing theory."¹³ Sections II, III and IV are three controlled behavioral experiments that develop and test key components of the theory. These experiments show how subtly different presentations of the identical DNA match statistics in a murder case affect (a) estimates of the chance that the suspect is the source of the matching DNA evidence, (b) estimates of the chance that the suspect is guilty, and (c) verdicts. One conclusion that emerges is that DNA match statistics that target the individual suspect and that are described as probabilities (e.g., "The probability that the suspect would match the blood drops if he were not their source is 0.1%") are more persuasive than statistical presentations that target a broad suspect population (e.g., all people in a large city) and that are framed as frequencies (e.g., "One in 1,000 people in Houston would also match the blood drops"). Section VI contains a fourth controlled experiment that identifies some limiting conditions on target and frame effects that may have practical importance in the courtroom. This section also includes discussion of why mock jurors may not appear to be persuaded by statistics as small as 1 in 1 billion. Section VI is a brief summary of the experimental findings and a discussion of experimental limitations. Section VII is a broader discussion of the psychology of DNA statistics. After examining the difference between legitimate and illegitimate statistical forms of DNA statistics, Section VII discusses how prosecutors and defense attorneys can use exemplar cueing to their advantage in the courtroom. Section VIII is a brief closing comment.

¹² Jury eligible subjects are U.S. citizens, at least 18 years old, who have not been convicted of a felony.

¹³ This theory was first identified in Jonathan J. Koehler and Laura Macchi, *Improving Jurors' Comprehension of Statistical DNA Evidence by Inducing an "Outside" Perspective*, Paper presented at the American Psychological

I. THE THEORY

A large body of research on statistical reasoning suggests that people have poor intuitions when it comes to reasoning with statistics in general¹⁴ and forensic science statistics in particular.¹⁵ Recent research with mock jurors indicates that the impressions left by DNA statistics vary as a function of perceived and actual error rates,¹⁶ expectations,¹⁷ and the mathematical form of the DNA statistic.¹⁸ Recently, Laura Macchi and I offered a theory that

Association convention, Chicago, IL (August 1997).

¹⁴ Terry Connolly, Hal R. Arkes, and Kenneth R. Hammond (eds.), *JUDGMENT AND DECISION MAKING: AN INTERDISCIPLINARY READER* (2000); Robin Hogarth *JUDGEMENT AND CHOICE* (2ND EDITION) (1991); Daniel Kahneman, Paul Slovic, and Amos Tversky (eds.) *JUDGMENT UNDER UNCERTAINTY: HEURISTIC AND BIASES* (1982); Michael J. Saks and Robert F. Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 *LAW AND SOCIETY REVIEW* 123 (1980-81). But see Leda Cosmides and John Tooby, *Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions From the Literature on Judgment Under Uncertainty*, 58 *COGNITION* 1 (1996).

¹⁵ David H. Kaye and Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence?*, 54 *J. ROYAL STAT. SOC. (Series A)* 75 (1991); Thomas D. Lyon and Jonathan J. Koehler, *The Relevance Ratio: Evaluating the Probative Value of Expert Testimony in Child Sexual Abuse Cases*, 82 *CORNELL L. REV.* 43 (1996) (documenting probabilistic reasoning errors made by courts in connection with evidence in child sexual abuse cases); William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence?*, 52 *LAW AND CONTEMP. PROBS.* 9 (1989). A series of studies on the use of non-DNA statistical evidence indicate that jurors are insufficiently sensitive to variations in the probativity of the match statistic. See David L. Faigman and A. J. Baglioni, Jr., *Bayes' Theorem in the Trial Process*, 12 *LAW AND HUMAN BEHAV.* 1 (1988); Jane Goodman, *PROBABILISTIC SCIENTIFIC EVIDENCE: JURORS' INFERENCES*, Doctoral Dissertation, University of Washington. (1986); Jane Goodman, *Jurors' Comprehension and Assessment of Probabilistic Evidence*, 16 *AMER. J. TRIAL ADVOCACY* 361 (1992); William C. Thompson and Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 *LAW AND HUMAN BEHAV.* 167 (1987); but see Brian C. Smith, Steven D. Penrod, A. L. Otto & R. C. Park, *Jurors' Use of Probabilistic Evidence*, 20 *LAW AND HUMAN BEHAV.* 49, 75-6 (1996) (mock jurors are sometimes sensitive to variations in both statistical and nonstatistical evidence).

¹⁶ Jonathan J. Koehler, Audrey Chia, and J. Sam Lindsey, *The Random Match Probability (RMP) in DNA Evidence: Irrelevant and Prejudicial?*, 35 *JURIM. J.* 201 (1995); Jason Schklar & Shari Diamond, *Juror Reactions to DNA Evidence: Errors and Expectancies*, 23 *LAW AND HUMAN BEHAV.* 159 (1999).

¹⁷ Schklar and Diamond, *supra*.

¹⁸ Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios and Error Rates*, 67 *U. COLORADO L. REV.* 859 (1996); Dale Nance, *An Empirical Examination of the Evidential Use of Likelihood Ratios*, Talk given at the International Conference on Forensic Statistics, (December 6, 1999); Franco Taroni and Colin G. G. Aitken, *Probabilistic Reasoning in the Law, Part 2: Assessment of Probabilities and*

may explain why some descriptions of DNA match statistics may have a larger impact on jurors than others.¹⁹ The theory may be described succinctly as follows: *the perceived probative value of a statistical DNA match (and, by extension, other forensic match evidence) depends on the ease with which triers of fact can imagine examples of others who would also match the DNA profile. When triers of fact find it hard to imagine examples of others who might match by chance, the evidence will be treated as compelling proof that the matching suspect is the source of the recovered DNA evidence. But when such matches are easier to imagine, the evidence will seem less compelling.* This theory is called "exemplar cueing theory." It is so named because it is based on an assumption that decision makers judge the probative value of a DNA match, in part, by the ease with which exemplars (examples) of others who might also match are cued (triggered) in their minds. When people find it hard to imagine such examples, it will seem reasonable to assume that the matching suspect is the source of the recovered DNA evidence. But when such examples are more easy to imagine (for whatever reason), the evidence will seem less compelling or, perhaps, insufficient. In the experiments below, I show that attorneys and experts may be able to influence whether or not jurors tend to think about others who might match simply by wording the DNA statistic in particular ways.²⁰

Explanation of the Value of Trace Evidence Other Than DNA, 38 SCIENCE AND JUSTICE 179 (1998a); Franco Taroni and Colin G. G. Aitken, *Probabilistic Reasoning in the Law, Part 1: Assessment of Probabilistic and Explanations Value of DNA Evidence*, 38 SCIENCE AND JUSTICE 165 (1998b).

¹⁹ Koehler and Macchi, 1997 Paper Presentation, *supra*.

²⁰ Data from some of the experiments were reported by the author in the following unpublished conference papers: *When are Jurors Impressed by DNA Match Probabilities?* Paper presented at the American Association for the Advancement of Science, Philadelphia, PA (Feb. 1998); *The Impact of Statistical Frame on Judgments About Probabilistic Evidence*, Paper presented at the American Psychology-Law Society Convention, Redondo Beach, CA (March 1998); *When are Jurors Impressed by Probabilistic Evidence?* Paper presented at the Judgment and Decision Making Society, Dallas, TX (Nov. 1998); *When are People Impressed by Probabilistic Evidence?* Paper presented at the European Science Foundation (Network on Human Reasoning and Decision Making), Milan, Italy (Feb. 1999); *Statistical Presentation Effects at Trial*, Paper presented at the International Conference on Forensic

Exemplar cueing theory has its roots in the “availability” heuristic²¹ from the judgment and decision making literature.²² Availability holds that people judge the frequency or probability of an event by the ease with which exemplars come to mind. The problem with judging frequency in this way is that exemplars may be readily available in mind for reasons that have nothing to do with their prevalence in the environment. Events may be available because they are widely publicized, interesting, unusual, vivid, personally relevant, etc. To give an example, people overestimate the number of people who die from lightning, fire, and firearms because these “flashy” causes of death are overrepresented in the news.²³ Unlike availability, exemplar cueing is not a heuristic that people use to estimate a frequency or a probability. Instead, it is a theory about how people determine the subjective weight to assign to the hypothesis that the matching suspect is the source of the recovered evidence.

According to exemplar cueing theory, triers of fact do not base their judgments about the probative value of a DNA match on the magnitude of the DNA match probability (where perceived probativity is inversely proportional to the size of the match probability). Instead, the perceived probativity of a DNA match is inversely proportional to the ease with which coincidental match “exemplars” are cued or present themselves in mind.²⁴ Thus, when triers of

Statistics, Raleigh, NC (Dec. 1999).

²¹ Amos Tversky and Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, 5 COGNITIVE PSYCH. 207 (1973); Russell B. Korobkin and Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CALIFORNIA L. REV. 1051, 1088 (2000)

²² Judgment and decision making research, which is frequently conducted by psychologists and economists, is the study of how people do and should make judgments and choices (particularly in probabilistic contexts).

²³ Baruch Fischhoff, Paul Slovic, and Sarah Lichtenstein, *Knowing With Certainty: The Appropriateness of Extreme Confidence*, 3 J. EXP. PSYCH: HUMAN PERCEPTION & PERF. 552, 554-564 (1977); Barbara Combs and Paul Slovic, *Newspaper Coverage of Causes of Death*, 56 JOURNALISM QUART. 837 (1979).

²⁴ Exemplar cueing is also consistent the more general notion that information is assigned weight in proportion to its “vividness” or ability to attract and hold our attention, Richard Nisbett and Lee Ross, HUMAN INFERENCE:

fact can easily think of instances in which such coincidences have occurred or can occur, they will find the evidence relatively less impressive.²⁵ If true, this could lead to the unusual situation in which objectively strong DNA evidence (as measured by its match probability) is accorded less value than objectively weaker DNA evidence when exemplars are cued for the strong evidence but not the weak evidence. As I discuss later, this is precisely what occurred in at least one study.

A. Target

One way to make people think about coincidental matches is to adjust the target of the DNA statistic away from the focal suspect and onto a larger reference population. This makes it easier for people to see that coincidental matches can occur, and that most of those who do match are not the source. Once the possibility of coincidental matches has been raised in this way, jurors may be less impressed with the DNA evidence. In contrast, when the target of the DNA match remains on the focal suspect, examples of coincidental matches are less likely to come to mind, and jurors will be more likely to treat small DNA matches as conclusive proof of identity.

To illustrate, suppose a juror hears that there is a 1 in 100,000 chance that a defendant

STRATEGIES AND SHORTCOMINGS OF SOCIAL JUDGMENT (1981) at 59.

²⁵ The notion that the impact of statistical evidence depends on examples that are cued in the decision maker's mind is related to the thesis that risk information that is presented in ways that promote frightening images is judged differently than mathematically identical information that does not promote such imagery. See Paul Slovic, John Monahan and Donald G. MacGregor, *Violence Risk Assessment and Risk Communication: The Effects of Using Actual Cases, Providing Instruction, and Employing Probability Versus Frequency*, 24 LAW AND HUMAN BEHAV. 271 (2000).

who is not the source of the genetic evidence would match by coincidence. The juror will probably not give much thought to the possibility that the match was a coincidence because 1 in 100,000 is very close to 0. This might constitute overwhelming proof of identify in the juror's mind. Now consider a juror who hears that 1 in every 100,000 people in, say, Houston who are not the source will match coincidentally. This juror may reason as follows: "If 1 in every 100,000 people in Houston match, then dozens of people from Houston would match, as would thousands of others throughout the U.S. For this reason, I do not find the evidence convincing as proof of identity."²⁶

I will refer to the difference between these two approaches as a difference in "target." DNA statistics may target a single defendant ("the frequency with which *the suspect* would match the blood drops if he were not their source is 1 in 100,000") or they may target a broader class of potential matchees ("One in 100,000 *people in Houston* would also match the blood drops"). I will refer to targets as single (suspect only) or multi (people in Houston).

B. Frame

Another factor that may influence whether jurors think about examples of coincidental

²⁶ At first blush, this line of reasoning *appears* to be similar to an approach that has been identified as the "defense attorney's fallacy." See William C. Thompson and Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutors' Fallacy and the Defense Attorney's Fallacy*, 11 LAW AND HUMAN BEHAV. 167, 171 (1987); David J. Balding & Peter Donnelly, *The Prosecutor's Fallacy and DNA Evidence*, CRIM. L. REV. 711, 711 (1994). However, there is an important difference. Whereas it *is* reasonable to point to the nonuniqueness of the evidence as evidence against its dispositive nature, it is not reasonable to dismiss the evidence as lacking probative value. The latter is the defense attorney's fallacy. For a discussion of the value of nonunique statistical evidence, see Michael Finkelstein and Will Fairly, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970).

matches is whether the DNA statistic is framed as a frequency (e.g., 1 in 100,000) or as a probability (e.g., .0.001%). The two frames are mathematically – but not psychologically – identical. Research on probabilistic reasoning shows that people reason differently with frequencies and probabilities.²⁷ Specifically, frequency frames encourage people to adopt a broad, “outside” view in which instant cases are viewed as a member of a larger class of events. This occurs because frequencies include a broad numerical reference class (e.g., 100,000) within which to consider the instant case. In contrast, probability frames do not include a reference class or broader context within which to think about the instant case. Therefore, probability frames induce people to adopt a narrower, “inside” view in which instant cases are thought of in isolation. Applying this idea to the DNA context, frequency framings of DNA statistics may induce jurors to think about that one coincidental match out of 100,000. In contrast, probability framings focus jurors’ attentions more narrowly on the suspect and discourages exemplar thoughts. If true, then we would expect jurors to be more impressed with and persuaded by DNA statistics that appear as probabilities (e.g., 0.001%) than those that appear as the equivalent

²⁷ Daniel Kahneman and Dan Lovallo, *Bold Forecasting and Timid Decisions: A Cognitive Perspective on Risk Taking*, 39 MAN. SCI. 17 (1993); Thomas Reeves and R. S. Lockhart, *Distributional Versus Singular Approaches to Probability and Errors in Probabilistic Reasoning*, 122 J. EXP. PSYCH: GENERAL 207 (1993); Amos Tversky and Daniel Kahneman, *Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment* 90 PSYCH. REV. 293 (1983).

One of the most striking examples of how people reason differently with probabilities and frequencies comes from Paul Slovic, John Monahan and Donald G. MacGregor, *supra* at 271. In this study, clinicians were asked to judge the risk that the patients would harm someone if discharged from a mental health facility. The results showed that the clinicians were less likely to recommend an immediate patient discharge than when they were given background information in relative frequency form (e.g., “of every 100 patients similar to Mr. Jones, 10 are estimated to commit an act of violence”), than when this information was provided in probability form (e.g., “patients similar to Mr. Jones are estimated to have a 10% probability of committing an act of violence”). See also Paul Slovic and John Monahan, *Probability, Danger and Coercion: A Study of Risk Perception and Decision Making in Mental Health Law*, 19 LAW AND HUMAN BEHAV. 49 (1995).

frequencies (1 in 100,000).²⁸

In short, exemplar cueing theory predicts that the way in which DNA statistics are presented influences whether or not jurors think about others who would also match by coincidence. If the DNA statistic focuses on the suspect (single target) or is framed as a probability (probability frame), then the evidence will be viewed as strong proof of identity. If the statistic promotes thoughts about others who might also match – either through the introduction of a broad target class (multi target) or the use of statistical frequencies (frequency frame), the DNA statistic will be less persuasive. Tests of this theory are presented below.

II. EXPERIMENT 1: CLINTON-LEWINSKY

I conducted an experiment involving 72 jury-eligible students (“mock jurors”) in January 1998 at the University of Texas to test whether there might be a relationship between ease of exemplar generation (as influenced by the presentation of statistical DNA evidence) and the persuasiveness of the evidence. At the time of the experiment, the United States was abuzz with an as yet unsubstantiated rumor that President Clinton had an affair with former White House intern Monica Lewinsky. Subjects were provided with the following background statement:

President Clinton has been accused of having a sexual relationship with former White House intern Monica Lewinsky. Mr. Clinton has denied the accusation. As part of his

²⁸ It is also possible that innumerate jurors will have trouble identifying the meaning of a probability that includes a series of zeroes to the right of the decimal point. They may not realize, for example, that 0.001% is mathematically equivalent to 1 in 100,000. Consequently, most of the studies reported here use percentage signs and slightly larger probabilities (e.g., 0.1%) to avoid confusing the problem of interest with an inability to make sense of decimals that include lots of zeroes.

investigation of these and other activities by Mr. Clinton, Special Prosecutor Kenneth Starr has collected clothing from Ms. Lewinsky, including a black cocktail dress.

Next, jurors were asked to assume that the following series of events takes place:

(1) Ms. Lewinsky continues to maintain her silence, (2) some genetic material (i.e., semen) is recovered from Ms. Lewinsky's cocktail dress, (3) a DNA expert reports that his tests could not rule out Mr. Clinton as a possible source of the recovered genetic material, (4) the expert also provides a statistic that describes the strength of the DNA match, and (5) Mr. Clinton's attorney responds by arguing that the DNA evidence is irrelevant because there are no corroborating eyewitnesses.

The case materials were identical for all subjects with the exception of the wording of the 4th event. Half of the subjects (selected at random) received the following "single target / probability frame" (s/p) wording: "the probability that Mr. Clinton would match the genetic material if he were not the source is 0.1%." The other half received the "multi target / frequency frame" (m/f) wording: "1 in 1,000 people in Washington who were not the source would also match the semen stain." Next, all subjects estimated the probability that President Clinton was the source of the DNA on the dress. I predicted that s/p subjects would be less sensitive to the possibility of coincidental matches than m/f subjects and would therefore assign higher estimates for the probability that President Clinton was the source of the DNA.

A. Results and Discussion

Jurors' responses were highly variable, ranging from 0% to 100%. This may have been

the result of using a sex scandal vignette in a politically charged environment. Nevertheless, a nonparametric analysis²⁹ revealed that s/p subjects thought it was significantly more likely that President Clinton was the source of the DNA than m/f subjects (82% vs. 60%).³⁰ If the probability estimates for both sets of subjects seem low in the face of a presumed DNA match, it may help to recall that this experiment was conducted: (a) when President Clinton's approval rating was at an all-time high (73%),³¹ (b) when 40% of Americans did not believe there was an affair,³² and (c) after several major news organizations (erroneously) reported that the FBI did not find evidence of semen on clothing taken from Ms. Lewinsky.³³

Another way of examining the results is in terms of the proportion of subjects who were certain or nearly certain that President Clinton was the source of the stain on the dress under the circumstances provided. The data showed that the proportion of subjects who were *at least 99% certain* that President Clinton was the source of the genetic material dropped from 28% among s/p jurors to 8% among m/f jurors.³⁴

²⁹ Nonparametric statistical tests are conservative methods for analyzing data. They are conservative in that they do not require any assumptions about the shape of the underlying parent population from which the sampled data are generated. The cost of this conservatism is a reduction in statistical power. This means that it is more difficult to achieve a statistically significant effect between groups with this test than with most other types of tests. See Michael O. Finkelstein and Bruce Levin, *STATISTICS FOR LAWYERS* (1990) at 309 (“Nonparametric tests are thus broadly applicable and valid for testing, relying only on the assumption of random sampling, and generally requiring only simple calculation.”)

³⁰ $Z=1.96$, $p=.05$. The nonparametric Z test compares the median (middle) estimates made by jurors in the two groups (s/p and m/f) to determine whether they are significantly different or not. When a Z test yields a p value $\leq .05$, the test is statistically significant. In the context of the present studies, the p value refers to the probability that the observed difference in the estimates made by jurors in the two groups could have arisen by chance alone.

³¹ CBS News, Evening News With Dan Rather (1/29/98).

³² ABC News, World News Tonight with Peter Jennings (1/30/98).

³³ CBS News, (1/29/98); World News Tonight, (1/30/98); see also S. Shepard, *That Mystery Dress: An Example of How Rumors Pass for Facts*, HOUSTON CHRON. at A6 (1/31/98).

³⁴ Wald χ^2 (1, $n=72$)=4.60, $p=.032$. A χ^2 (“chi-square”) statistic is a widely used test statistic for comparing the independence of observations collected across groups. See Finkelstein and Levin, *supra* at 159 (“... the chi-squared statistic is a reasonable test statistic for the disparity between observed and expected frequencies in a 2 X 2 table.”)

Obviously there are many variables at work in a loaded context such as this. Politics, the influence of the news media, prior beliefs about the credibility of Clinton and Lewinsky, prior beliefs about the nature of the relationship between Clinton and Lewinsky, etc. Moreover, this experiment does not indicate whether the difference in target (single or multi), frame (probability or frequency) or both *caused* the observed differences between groups. In order to disentangle the central effects and put exemplar cueing theory to a more critical test, Experiment 2 was conducted using DNA matches in a hypothetical criminal context.

III. EXPERIMENT 2: TARGET AND FRAME EFFECTS

Imagine the following scenario: A masked man bursts into a Houston hardware store and announces his intent to rob it. Shortly thereafter, the owner of the store tries to overtake the would-be robber and strikes him on the head with a hammer. However, the robber shoots and kills the owner before fleeing from the store. Blood that does not belong to the storeowner is recovered at the crime scene and provides a DNA match with a suspect. At trial, a DNA expert testifies about the match and offers some probabilities that are intended to help jurors understand the probative value of the DNA evidence. Other than the DNA evidence, the case against the suspect is weak.

Ninety mock jurors³⁵ read a case that included facts similar to this one.³⁶ As in the

The larger the disparity, the greater the value of the statistic.”) As in the Z test reported previously, when a χ^2 test yields a p value $\leq .05$, the differences observed across groups is statistically significant.

³⁵ The mock jurors in the experiments were full-time students enrolled in classes at the University of Texas at Austin. Approximately 10% identified themselves as non-U.S. citizens and their data were excluded. See Section

Clinton-Lewinsky study, the information provided was identical for all jurors with the exception of a single sentence that described the DNA match statistic. The DNA match statistic, which was fixed at 0.001 (one in one thousand), was described in one of four mathematically equivalent ways: (1) single target / probability frame (s/p): "The probability that the suspect would match the blood drops if he were not their source is 0.1%"; (2) single target / frequency frame (s/f): "The frequency with which the suspect would match the blood drops if he were not their source is 1 in 1,000"; (3) multi target / probability frame (m/p): "0.1% of the people in Houston would also match the blood drops"; (4) multi target / frequency frame (m/f): "One in 1,000 people in Houston would also match the blood drops." The mock jurors were assigned at random to one of these four groups (22-24 jurors per group) that differed only in terms of the presentation of the

VI for a brief discussion on the limitations – both real and imagined – of using jury-eligible student-subjects in mock trial experiments.

³⁶ A sample experimental case is as follows:

"In the case of State of Texas v. Nethers, Steven Nethers was accused of murdering Richard Oden during an attempted robbery of a hardware store owned by Mr. Oden in Houston, Texas.

According to reliable eyewitness accounts, the perpetrator entered Mr. Oden's hardware store at approximately 4:30 p.m. on November 2, 1997 wearing a Halloween-type of mask, and waving a small caliber handgun. The perpetrator approached Mr. Oden (who was behind the cash register) and said, "Open it fast or you're a dead man." According to the eyewitnesses, when the perpetrator turned his head to survey the store, Mr. Oden grabbed a hammer from the counter and smashed the perpetrator on head with a single blow. The perpetrator fired a single shot into Mr. Oden's chest and fled the store. Mr. Oden died shortly thereafter in a Houston hospital.

During an investigation of the hardware store crime scene, Houston police identified and recovered several moist blood drops from the path that was taken by the perpetrator as he fled the store. These drops were subjected to a form of DNA analysis called PCR testing. The PCR tests revealed the blood to be of a type known as "2, 3." Because this blood type was different from Mr. Oden's blood type, police believed that the recovered blood drops came from the bleeding head of the robber.

During routine interviews of people who live in neighborhood, the police identified several potential suspects. All of these individuals agreed to provide blood samples to police for comparison with blood that was recovered from the crime scene. One of the suspects, Mr. Steven Nethers, matched the 2, 3 blood type and was arrested for the murder.

At trial, the prosecution alleged that the blood analysis demonstrated Mr. Nethers was the source of the wet blood drops and that he was therefore guilty of attempted robbery and murder.

A DNA expert testified that his tests could not rule out Mr. Nethers as a possible source of the blood drops. He also testified that 1 in 1,000 people in Houston who are not the source would nonetheless match the blood drops.

The defense argued that the blood evidence is irrelevant because there was no direct evidence (such as

DNA match statistic. After reading the case, subjects (a) estimated the probability that the suspect was the source of the recovered DNA trace, “p(source)”, (b) estimated the probability that the suspect was guilty of the murder, “p(guilt)”, and (c) provided a verdict (guilty, not guilty).

Exemplar cueing theory predicts that jurors exposed to the frequency frame and multi target presentations of the DNA statistic will be more likely to think about examples of coincidental matches than jurors exposed to the probability frame and single target presentations. Consequently, I predicted that the frequency frame and multi target groups would not be as impressed by the DNA evidence and would assign lower values to p(source) and p(guilt), and return fewer guilty verdicts than the probability frame and single target groups.

A. Results and Discussion

As predicted, the results showed that jurors' p(source) and p(guilt) estimates varied depending on how the DNA match statistic was targeted and framed. Mock jurors in the single target groups concluded that the suspect was significantly more likely to be the source and more likely to be guilty than did jurors in the multi target groups (p(source) averages: 79% vs. 61%; p(guilt) averages: 78% vs. 62%).³⁷ Likewise, mock jurors in the probability frame groups concluded that the suspect was significantly more likely to be the source and more likely to be

eyewitness identifications) that linked Mr. Nethers to these crimes."

³⁷ A multivariate analysis of variance (MANOVA), was performed on jurors' p(source) and p(guilt) estimates. This analysis is commonly used to examine the effects of independent variables (such as target and frame) on more than one dependent variable (such as p(source) and p(guilt)). It produces an F test statistic and p values. As in the statistical tests discussed earlier, when an F test yields a p value $\leq .05$, the differences observed across groups are

guilty than did jurors in the frequency frame groups (p(source) averages: 80% vs. 60%; p(guilt) averages: 74% vs. 66%).³⁸ Jurors in the single target groups tended to return more guilty verdicts than jurors in the multi target groups (40% vs. 29%), although this difference fell short of statistical significance. Similarly, probability frame jurors tended to convict more often than frequency frame jurors (36% vs. 32%),³⁹ but here too the difference did not reach statistical significance.⁴⁰

The proportion of "extreme responses" ($\leq 1\%$ or $\geq 99\%$) in the s/p and m/f groups provides insight into the influence these subtle presentation manipulations have on people's beliefs about the value of the statistical evidence. Figure 1 shows the proportion of extreme p(source) responses made by jurors in the s/p and m/f groups.⁴¹ Most s/p jurors (63%) concluded that the suspect was at least 99% likely to be the source of the blood. Only 8% of the jurors concluded there was a 1% chance or less that the suspect was the source of the blood. In

statistically significant. Here, the MANOVA revealed significant effects for target, $F(2,84)=3.02$, $p=.05$.

³⁸ $F(2,84)=3.96$, $p=.02$. In general, p(guilt) estimates mirrored p(source) estimates but were somewhat lower in all groups. This is understandable when we consider that, for many crimes, one may be the source of recovered DNA evidence yet may not be guilty of the crime. For example, the sample may have been left innocently at an earlier time. We might expect the difference between p(source) and p(guilt) estimates to be smaller in rape cases where the contested issue is perpetrator identity.

³⁹ Although the verdict data fell short of statistical significance, the results are consistent with unpublished data that employed a different incidence rate. Andrea House conducted an experiment with 124 mock jurors who heard a DNA match statistics either as a frequency (1 in 50,000) or a probability (.002%). Those who heard the probability evidence were nearly 50% more likely to return a guilty verdict than those who heard the evidence presented as a frequency, (probability frame=63.3% vs. frequency frame=43.7%, $\chi^2(1)=4.83$, $p < .05$). Andrea House, *Methods of Presentation of Expert Statistical Testimony and its Influence on Jury Decision-Making*, Unpublished Manuscript (December 2, 1996).

⁴⁰ One reason that statistical significance was not achieved on the verdicts is that the statistical power associated with detecting binary verdict differences is lower than that associated with detecting differences on continuous probability measures such as p(source) and p(guilt). William L Hays, *STATISTICS FOR PSYCHOLOGISTS* (1963) at 619.

⁴¹ The proportions of extreme p(source) responses in the s/f and m/p groups were in between those of the s/p and m/f groups. For the s/f group, 17% of the p(source) responses were $\leq 1\%$ and 38% were $\geq 99\%$. For the m/p group, 10% of the p(source) responses were $\leq 1\%$ and 25% were $\geq 99\%$.

contrast, only 14% of the m/f jurors concluded that the suspect was at least 99% likely to be the source of the DNA. Incredibly, 32% of m/f jurors concluded that there was less than a 1% chance that the suspect was the source of the blood! In short, whereas more than one-half of the jurors in the s/p group were quite certain that the suspect was the source of the DNA evidence, about one-third of the jurors in m/f group were quite certain that the suspect was *not* the source. This disparity emerged even though the cases presented were identical across groups except for the wording of a single sentence that described identical DNA match probabilities.

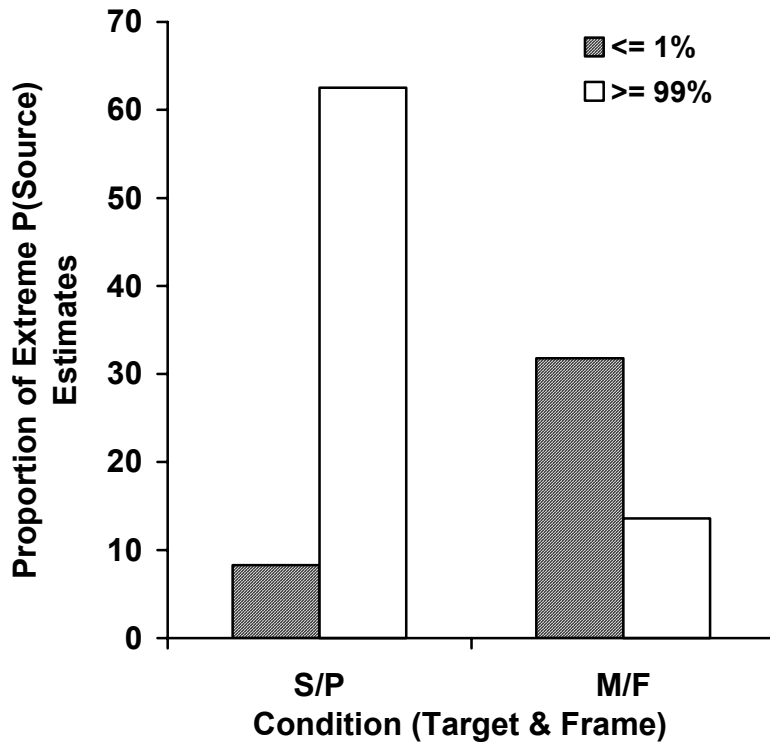


Figure 1. Proportion of extreme ($\leq 1\%$ and $\geq 99\%$) p(source) estimates by mock jurors (DNA incidence rate=.001). S/P = single target, probability frame; M/F = multi target, frequency

frame.

The major finding here is that jurors were substantially less impressed with the DNA evidence when it was presented in m/f form than in s/p form. Presumably, m/f jurors, unlike s/p jurors, were cued to imagine others who might match by coincidence and assigned relatively lower $p(\text{source})$ and $p(\text{guilt})$ probabilities in response.⁴² Verdict, a binary variable, was a less sensitive measure and therefore did not differ significantly across most groups. Nevertheless, if future studies confirm the trends detected here, the legal system must take seriously the idea that the way in which a match statistic is worded by an expert or attorney can affect the way a juror thinks about the value of that evidence.

IV. EXPERIMENT 3: DUAL PERSPECTIVE

In an actual trial, the defense and prosecution may try describe DNA evidence in ways that are most favorable to their favored position.⁴³ In cases that include a DNA match, the prosecution may wish to present the evidence in terms of probability frames and single targets while the defense recharacterizes that evidence in frequency frames and multi targets. This

⁴² Anecdotal support for this explanation also appeared in written comments made by some mock jurors. A typical comment by a juror in the exemplar-conducive groups (multi target or frequency frame) was: “If 1 in 1000 people would match this blood, it is very likely that [the defendant] is not the one who murdered [the victim]. In contrast, a typical comment by a juror in the single target and probability frame groups was: “The odds that there is another suspect that has his same blood match are nearly zero.”

⁴³ The operative word in the sentence above is “try.” Trial transcripts are replete with examples of *defense attorneys* who mischaracterize DNA evidence in ways that make it appear to be stronger evidence than it really is. In a capital murder case, defendant’s counsel stated that blood scrapings recovered from a bottle in the possession of the defendant were “95% probably [the victim’s] blood,” *Dowhitt v. State* 931 S.W. 2d 244 (Tex.Cr.App. 1996), as reported in Stuart Paul O’Brien, *Dangerous Numbers for the Accused?* Unpublished Manuscript (April, 1999), p.

prompts the empirical question of how persuasive the DNA evidence is for mock jurors who are presented with *both* target and both frame perspectives.

A new set of 227 jury-eligible University of Texas subjects participated in this study for credit in an introductory business law class. Subjects were assigned at random to one of six experimental groups and provided with a hypothetical case and set of questions that were largely the same as those given to mock jurors in Experiment 2. Indeed, four of the six groups were identical to those used in Experiment 2 (25-30 jurors per group).⁴⁴ The two remaining groups were dual perspective groups (55-58 per group). These included both an *s/p* and an *m/f* format for the DNA statistic. One of the dual perspective groups used an *s/p-m/f* order and the other used an *m/f-s/p* order.

How will the dual presentation jurors respond? One possibility is that dual presentation jurors will respond similarly to *m/f* jurors. This is because both types of jurors will be exposed to a presentation form that induces them to think about others who might match. By this reasoning, the dual presentation jurors and *m/f* jurors will be less impressed with the evidence than *s/p* jurors. A second possibility is that dual presentation jurors will anchor on their initial impressions of evidence strength and pay less attention to the recharacterized statistical evidence. This is consistent with the anchoring and adjustment theory of decision-making⁴⁵ and the primacy effects predicted by the Belief Adjustment Model for single-judgment decision

24. For other examples, see Koehler, 1993, *supra* at 35.

⁴⁴ Perhaps the best way to determine whether a scientific result is reliable is to replicate it. Julian L. Simon and Paul Bernstein, *BASIC RESEARCH METHODS IN SOCIAL SCIENCE* (3rd Ed.) at 134 (1985).

⁴⁵ Amos Tversky and Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, in *JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES* 3, 14-18 (Daniel Kahneman, Paul Slovic and Amos Tversky, eds., 1982).

making tasks.⁴⁶ A third possibility is that dual presentation jurors will adopt a little of each perspective and end up with views on the critical questions between those of the s/p and m/f groups.

A. Results and Discussion

The data from the two dual presentation conditions did not differ as a function of the order in which the s/p and m/f perspectives appeared. Therefore, the data from these groups (groups five and six) were collapsed into a single group for all other analyses.

The data revealed that mock jurors' p(source) and p(guilt) estimates varied across the five groups.⁴⁷ S/p jurors were most impressed with the evidence (p(source) average=77%, p(guilt) average=68%) and m/f jurors were least impressed (p(source) average=38%, p(guilt) average=38%).⁴⁸ Dual presentation jurors gave responses that fell in between (p(source) average=50%, p(guilt) average=46%). On the verdict measure, dual presentation jurors were about as willing to convict as s/p jurors (22% vs. 26%). However, m/f jurors rarely convicted (3%).

These data are important for two reasons. First, they replicate the presentation effects in Experiment 2, which showed that the way statistical DNA evidence is targeted and framed affects the persuasive value of that evidence. Second, the data suggest that introduction of a dual

⁴⁶ Robin M. Hogarth and Hillel J. Einhorn, H. J., *Order effects in Belief Updating: The Belief-adjustment Model*, 24 COGNITIVE PSYCH. 1 (1992).

⁴⁷ $F(8,440)=2.90$, $p=.004$

⁴⁸ Responses by jurors in the s/f and m/p groups fell in between those of s/f and m/p jurors on all of the questions. The p(source), p(guilt), and proportion of guilty verdicts by the s/f group were: 57%, 49% and 24%, respectively.

perspective on the DNA statistic changes the way jurors think about DNA evidence – to a point. Some of the judgments and decisions made by jurors who received *both* s/p and m/f statistical presentations differed from those of jurors who received the s/p or m/f presentation alone. This suggests that introduction of the 2nd presentation – particularly an s/p presentation – changed how jurors thought about the evidence

Those who received the m/f perspective in addition to the s/p perspective gave lower p(source) and p(guilt) estimates than those who received only the s/p perspective. This may have occurred because the m/f perspective reminded the jurors that others would also match. Nevertheless, these dual presentation jurors convicted about as often as the s/p jurors. Thus, although the m/f perspective weakened their belief about the defendant's guilt, it apparently was not enough (in this scenario) to change their minds about the verdict.

Those who received the m/f presentation alone were notably less persuaded by the evidence than were those who received the s/p perspective in addition to the m/f perspective. Apparently, jurors in this group focused on the possibility that others would also match, and few were impressed with the evidence under these circumstances. Importantly, the conviction rate increased notably (from 3% to 22%) once the s/p perspective was introduced. Apparently, introduction of the s/p perspective can refocus jurors' attention on the unlikelihood that this particular defendant would match if he were not the source even among jurors who previously were not persuaded by the evidence.

The corresponding figures for the m/p group were: 55%, 57%, and 24%.

V. EXPERIMENT 4: SMALLER MATCH PROBABILITIES

The first three experiments used DNA matches that had an incidence rate of one in one thousand. But DNA match probabilities that find their way into the courtroom are often as small as one in millions or billions.⁴⁹ Will the way in which the statistics are presented really matter for such small match probabilities? The answer provided by exemplar cueing theory is that it depends on whether jurors are able to generate coincidental match exemplars in some presentations but not others. If match exemplars cannot be generated in *any* presentation group, then we would expect that the effects for target and frame observed in previous experiments to disappear.

With this in mind, Experiment 4 examined the effects of target and frame at very low DNA incidence rates (one in one million, one in one billion) on 282 jury-eligible students. The case materials were identical to those used in Experiments 2 and 3. Subjects were assigned at random to one of 8 groups. The groups consisted of all possible combinations of target (single, multi), frame (probability, frequency) and incidence rate (1 in 1 million, 1 in 1 billion) groups.

I predicted that the target and frame effects would be smaller or even nonexistent at these very low incidence rates. This prediction followed from the idea that exemplars may be hard to generate in any of the target and frame groups (including the m/f group) for very small incidence rates.

⁴⁹ Jonathan J. Koehler, *One in Millions, Billions and Trillions: Lessons from People v. Collins(1968) for People v. Simpson (1995)*, 47 J. LEGAL EDUC. 214 (1997).

A. Results and Discussion

As expected, the data in Experiment 4 showed that the target and frame effects observed earlier depend on incidence rate levels. At the 1 in 1,000,000 incidence rate, there were small differences between the single target and the multi target on $p(\text{source})$ estimates (81% vs. 76%). There were also small differences between the probability frame and the frequency frame on $p(\text{source})$ estimates (82% vs. 74%). Similarly, there were small differences in $P(\text{Guilt})$ estimates and verdict across the different target and frame groups in the expected direction. Apparently, then, the form of the DNA statistic still mattered a bit even at an incidence rate as small as 1 in 1,000,000. This suggests that some people may think about match exemplars even when the incidence rate is very small.⁵⁰

At the 1 in 1,000,000,000 (1 in one billion) incidence rate, all differences between the different target and frame groups disappeared. For example, there were no differences between the single target and the multi target on $p(\text{source})$ estimates (83% vs. 83%) or between the probability frame and the frequency frame on $p(\text{source})$ estimates (84% vs. 82%). As the exemplar cueing theory acknowledges, the notion that a 1 in 1,000,000,000 match could be coincidental is, for most people, farfetched because it is hard to imagine examples of others in *any* reference population (except, perhaps, the entire world) that would match by coincidence. It appears, then, that exemplar reasoning requires a reference space that is sufficiently large that exemplars may be identified.

⁵⁰ This is not too surprising when one considers that several people in the large U. S. cities would share a genetic profile that has an incidence rate as small as 1 in 1,000,000.

1. *One In One Billion: Not Persuasive?*

At first blush, one surprising finding from Experiment 4 is that jurors were not utterly convinced by the DNA statistics, even when the match statistic was 1 in 1 billion. These jurors assigned average source and guilt probabilities of 83% and 77% respectively. Why weren't they absolutely convinced by the DNA evidence?

One reason is that the hypothetical case used in the experiment was weaker than the typical DNA case. Aside from evidence of a DNA match, the prosecution did not offer additional evidence against the defendant. I intentionally used such a case in order to avoid “ceiling effects” in which jurors from all experimental groups became firmly convinced of the defendant’s guilt. Ceiling effects caused by overwhelming evidence of guilt could mask the subtle effects predicted by exemplar cueing theory.

Second, the reluctance of some jurors to assign extremely high $p(\text{source})$ and $p(\text{guilt})$ values in mock DNA cases is now a fairly consistent finding in the mock juror literature.⁵¹ This suggests that mock jurors can only be so impressed with DNA evidence. As the DNA match probability gets smaller, the probative value of the DNA evidence (holding aside error rate considerations) gets larger. However, the weight that mock jurors assign to such evidence may be limited for several reasons. One possibility is that jurors understand that the value of DNA

⁵¹ Koehler, Chia, and Lindsey, *supra*; Nance, *supra*; Schklar and Diamond, *supra*; Smith, Penrod, Otto and Park, *supra*; Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 J. PERSONALITY & SOCIAL PSYCH. 739 (1992).

evidence is limited by the chance that the reported match is erroneous.⁵² Support for this is found in a recent study that reported that the average juror assumes the chance of an erroneous DNA match call is on the order of 1 in 15.⁵³ Another possibility is that jurors regard a DNA match to be a form of “naked statistical evidence” and are unwilling to assign extreme weights to such evidence without corroborating evidence.⁵⁴

Finally, the summary data showing jurors' reluctance to assign extremely high source and guilt estimates when presented with 1 in 1 billion DNA match statistics may be misleading. The 83% source probability and the 77% guilt probability values are averages (means). As such, they include responses from all jurors, including a few who gave low source and guilt probability estimates. A few low estimates can move an average (i.e., an arithmetic mean) away from the median values thereby giving a misleading impression of the "typical" juror. In fact, the median p(source) and p(guilt) estimates for the jurors in the one in one billion group were 98% and 90% respectively.

Nevertheless, the minority jurors should not be ignored. Six percent of jurors who were provided with the 1 in 1 billion DNA match statistic (9 of 147) assigned source probabilities \leq 10%, and 3% (5 out of 147) assigned a source probability \leq 1%. Eight percent of the jurors (12 out of 147) assigned guilt probabilities \leq 10% and 4% (6 out of 147) assigned a guilt probability \leq 1%. These jurors were almost certainly confused. Apparently they thought that the extremely

⁵² False match errors due to sample mix-ups, contamination, etc. limit the probative value of reported DNA matches. Koehler, Chia, and Lindsey, *supra* at 216 ("the evidentiary significance of a reported DNA match is often informed more by the possibility of laboratory error than by the possibility of a coincidental match"); Lempert, 1991, *supra*; Richard O. Lempert, *Why Do Jury Research?*, 242 INSIDE THE JUROR (Reid Hastie ed, 1991).

⁵³ Schklar and Diamond, *supra* at 178.

⁵⁴ David H. Kaye, *DNA Evidence: Probability, Population Genetics, And the Courts*, 7 HARV. J. LAW & TECH. 101 (1993); Wells, *supra*.

small DNA match probability made it less (rather than more) likely that the suspect was the source. If such confused jurors appear at rates of roughly 6% or 8% on actual juries, the odds are that a jury of 12 will have one or more such individuals.⁵⁵ This could be the difference between a conviction and a hung jury.

VI. THE EXPERIMENTS: A SUMMARY AND LIMITATIONS

A. Summary

This paper represents the first attempt to offer and test a theory for predicting when legal decision makers will and will not be impressed with low probability DNA match statistics. The results dramatically illustrate that the persuasive power of DNA evidence may depend, in part, on how the match statistics are presented and framed. Experiments 1, 2 and 3 showed that when evidence is presented in ways that prompt triers of fact to think about examples in which coincidental DNA matches might occur, the evidence loses some of its persuasive force. But when such examples are not readily called to mind, the evidence has more persuasive force. Experiment 4 provided an important limiting condition for target and frame effects. When DNA incidence rates are so small that exemplar generation is difficult under any target and frame formulation, the way in which the statistics are presented is less significant.

⁵⁵ If 6-8% of people are “confused,” then the probability of at least one confused juror appearing on a 12-person jury is $1-(.94)^{12}=.52$ when 6% of jurors are confused, and $1-(.92)^{12}=.63$ when 8% of jurors are confused.

B. Limitations

All experiments have limitations and those presented here are not exceptions. Below, I briefly address two limitations – one imagined and one real. I also review data one small, but telling, study.

1. “*Students Aren’t Jurors*”: *Not A Fatal Flaw*

The subjects in these experiments were jury-eligible University students rather than either a broader cross-section of people or a sample of empanelled jurors. Although University students are younger and more intelligent than the average juror,⁵⁶ a large empirical literature provides little reason to believe that patterns of data obtained from student-subjects fail to generalize to the jury population.⁵⁷ Moreover, in a previous DNA study, patterns of data obtained from student-subjects did not differ meaningfully from patterns obtained from empanelled jurors.⁵⁸ Thus, the statistical confusion that some of our subjects exhibited in Experiment 4 probably exists among actual jurors. Stuart O'Brien finds support for this proposition in an unpublished study.

⁵⁶ The business students used here are probably more quantitatively savvy than the average juror. Not only is the curriculum quantitative, but the undergraduate business program at the University of Texas at Austin is ranked 5th in the nation by U.S. News and World Report, <http://www.usnews.com/usnews/edu/college/rankings/cat12rnk.htm> (2000).

⁵⁷ Brian H. Bornstein, *The Ecological Validity of Jury Simulations: Is the Jury Still Out?*, 23 LAW AND HUMAN BEHAV. 75 (1999).

⁵⁸ Koehler, Chia, and Lindsey, *supra* at 214.

a. *Four confused jurors – The O'Brien study*

Stuart O'Brien gained access to four actual jurors in a Texas capital murder case several years after the jury convicted the defendant.⁵⁹ The conviction was partially based on a PCR DNA blood match between the victim and a spot on the clothing owned by the defendant. The test revealed that the blood matched the blood type of the victim. The test gave a DNA match statistic of 1 in 20 (5%).

O'Brien posed a hypothetical to the four former jurors that was similar to the hypothetical case used in the experiments reported here. He asked them to imagine a murder case in which an expert testified that (a) DNA evidence recovered from clothing worn by a suspect matched the victim, and (b) the frequency of this DNA profile in the general population is 1 in 100. O'Brien then presented the former jurors with a series of statements related to the meaning of the 1 in 100 statistic and asked them to indicate whether each was true or false.

In general, their performance was abysmal. One of the jurors accepted as true the misstatement that a "1 in 100 frequency indicates that there is a 99% chance that the victim is the source of the evidence." This is the source probability error.⁶⁰ The error consists in equating the frequency of a profile (e.g., a DNA profile) with the probability that a person who matches the profile is not the source of that profile. Thus, when judges, experts, and attorneys claim that a DNA match probability of, say, 1 in 1,000,000 means that there is only one chance in a million that the suspect is not the source of the recovered sample, they have committed the source

⁵⁹ Dowhitt v. State 931 S.W. 2d 244 (Tex.Cr.App. 1996).

⁶⁰ Jonathan J. Koehler, 1993 *supra* at 26.

probability error.⁶¹

The statistical error committed by the other three jurors was worse. They concluded that the 1 in 100 statistic indicated that there was only a 1% chance that the blood belonged to the victim. This is false. By equating the profile frequency with the source probability, these jurors turned the notion of probative value on its head in much the same way as did some of the mock jurors in Experiment 4. If the profile frequency actually did equal the source probability, then an extremely rare blood match (e.g., 1 in 1 million) would be *less* probative than an extremely common blood match (e.g. 4 in 5) because there would be an 80% that the blood belonged to the victim in the latter case but less than a 1% chance in the former case. This is obviously wrong. Apparently, then, service on a jury in which DNA evidence plays an important role provides little reason to believe that people who reason badly about DNA statistics in a laboratory setting will improve in a courtroom setting.

2. “Real Cases Are More Complex”: A Better Criticism

A more useful criticism of the present studies is that the case materials were not as rich as those that jurors have in an actual trial. This was intentional. In a controlled scientific study, it is important to minimize interference from other variables to determine whether there is a causal

⁶¹ Id at 27; David H. Kaye and George F. Sensabaugh, G. F. *Reference Guide on DNA Evidence* 485, 544 REFERENCE MANUAL ON SCIENTIFIC EVIDENCE, 2ND ED (Federal Judicial Center, 2000). The source probability error is closely related to the “prosecutor’s fallacy” described in Thompson and Schumann, *supra*, at 171 . The fallacy consists in equating the frequency of the match profile with the probability that the suspect did not commit the crime he/she is accused of committing.

relationship among the variables of interest.⁶² However, such causal control comes at a price, namely, a reduction in realism. In actual DNA cases, jurors are probably influenced by numerous factors other than the way in which the DNA evidence is presented. For example, real jurors hear extended opening and arguments, observe direct and cross-examinations of witnesses, deliberate as a group, and consider the consequences of their verdicts.

Future researchers may wish to trade off some of the rigor and experimental control of the present studies in favor of increased realism. For example, they might wish to compare outcomes in cases where the DNA statistics were presented in s/p form with those in which the statistics were presented in m/f form. Although causal relationships may not be inferred from such a correlational study, these data could bolster or challenge the exemplar thesis.

VII. GENERAL DISCUSSION: THE PSYCHOLOGY OF DNA NUMBERS

Some may find it alarming that the wording of a statistic can affect the chance of a defendant's guilt in the eyes of a juror. Yet this observation is consistent with a large body of research on the psychology of numbers. Much of this research indicates that people think heuristically rather than probabilistically. That is, when presented with quantitative information, we do not perform algebraic computations and arrive at solutions by using tenets of logic and probability theory.⁶³ Instead, we evaluate quantitative evidence via mental shortcuts and other

⁶² William S. Ray, AN INTRODUCTION TO EXPERIMENTAL DESIGN (1960) at 6-14.

⁶³ Daniel Kahneman, Paul Slovic and Amos Tversky in JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES (Eds. 1982) Cambridge: Cambridge University Press.

rules of thumb.⁶⁴ In the case of DNA evidence, the ease with which we can imagine scenarios or examples in which someone other than the matching suspect is the source of a sample may be the heuristic of choice.

A. Statistical Form vs. Statistical Fallacy: An Important Distinction

As noted earlier, all of the target and frame presentations used in the experiments reported here are legitimate, mathematically comparable ways of presenting a DNA match statistic. In contrast, some reformulations of statistical evidence are fallacious. For example, it is not legitimate to present the DNA match statistic as a statement about (a) the probability of guilt or innocence,⁶⁵ (b) the source probability,⁶⁶ or (c) the probability that another person would match.⁶⁷ These fallacious reformulations implicitly assume knowledge about the strength of the non-genetic evidence against the matching suspect, or the size of the reference population (i.e., the number of others who could be the source).

⁶⁴ Legal scholars recently took note of this literature and reported that heuristic ways of thinking about probabilistic matters are common in legal settings. See Richard Birke and Craig R. Fox, *Psychological Principles in Negotiating Civil Settlements*, 4 HARV. NEGOTIATION L. REV. 1 (1999); Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471 (1998); Mark Kelman, Yuval Rottenstreich and Amos Tversky, *Context Dependence in Legal Judgment*, 25 J. OF LEGAL STUD. 287 (1996); Russell Korobkin, *The Status Quo Bias and Contract Default Rules*, 83 CORNELL L. REV. 608 (1998); Russell Korobkin and Chris Guthrie, *Psychological Barriers to Litigation Settlement: An Experimental Approach*, 93 MICH. L. REV. 107 (1994); Russell Korobkin and Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 89 CAL. L. REV. (2000); Donald Langevoort, *Behavioral Theories of Judgment and Decision Making in Legal Scholarship*, 51 VAND. L. REV. 1499 (1998); Jeffrey J. Rachlinski, *Gains, Losses, and the Psychology of Litigation*, 70 S. CAL L. REV. 113 (1996); Jeffrey Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571 (1998); Dan Simon, *A Psychological Model of Judicial Decision Making*, 30 RUTGERS L. J. (1998); Cass R. Sunstein, *Behavioral Analysis of Law*, 64 U. OF CHI. L. REV. 1175 (1997); Cass R. Sunstein (ed.) BEHAVIORAL LAW AND ECONOMICS (2000).

⁶⁵ Thompson and Schumann, *supra*; Balding and Donnelly, *supra*; David J. Balding and Peter Donnelly, *Inference in Forensic Identification*, 158 J. ROYAL STAT. SOC., SERIES A 21, 35 (1995).

⁶⁶ Koehler, 1993, *supra* at 27.

Admittedly, the multi target presentation (which identifies the proportion of others in a population who would match) bears a superficial resemblance to the "defense attorney's fallacy" in which statistical evidence is undervalued or dismissed on grounds that innocent others would also be falsely implicated by the match.⁶⁸ The difference between the multi target presentation technique and the defense attorney's fallacy, however, is that the latter includes a faulty interpretation of the fact that the statistical profile is not unique. Other things being equal, nonunique evidence *is* less probative than unique evidence. However, nonunique evidence may still be extremely probative.⁶⁹ The multi target presentation encourages an awareness of the nonuniqueness of the evidence, but it does not provide an interpretation (faulty or otherwise) of that evidence.

B. Which Statistical Form is "Best?"

⁶⁷ Koehler, 1993 *supra* at 34.

⁶⁸ Thompson and Schumann, *supra*. See also fn 26.

⁶⁹ Many scholars agree that the probativity of statistical evidence can be represented in terms of a likelihood ratio, Kaye, *supra* at 161, Richard D. Friedman, *A Presumption of Innocence, Not of Even Odds*, 52 STAN. L. REV. 873, 875 (2000); Lyon and Koehler, *supra* at 47; Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 STAN. L. REV. 1477, 1486 (2000). Specifically, probativity of evidence (E) with respect to guilt (G) is given by this ratio: $P(E|G) / P(E|-G)$, Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1026 (1977) ("evidence is logically relevant only when the probability of finding that evidence given the truth of some hypothesis at issue in the case differs from the probability of finding the same evidence given the falsity of the hypothesis at issue"). Non-unique evidence is probative in favor of the Guilt hypothesis to the extent that the denominator of the likelihood ratio, $P(E|-G)$, is *small* relative to the numerator of the likelihood ratio. In cases involving DNA evidence, the denominator of the likelihood ratio is highly correlated with (and sometimes identical to) the DNA match statistic. Thus, as the incidence rate given by a nonunique DNA match statistic becomes small, the likelihood ratio becomes large, and the DNA evidence becomes increasingly probative.

Whether likelihood ratios are easily understood by decision makers is another story altogether. See Koehler, 1996, *supra* at 876-883 (showing that experts and mock jurors confuse likelihood ratios with posterior odds ratios); *State v. Garcia*, 3 P. 3d 999, 1002, n. 2 (Ariz. 1999) (indicating likelihood ratio confusion among the prosecutor, defense attorney, presiding judge, and jurors in a rape case involving DNA evidence); Kaye and Koehler, *supra* at 76 (reviewing studies that show an inverse error); Kaye, 1993, *supra* at 163 (noting that a jury

Although the different ways of presenting the DNA match statistic offered here are legitimate, they may not be equally likely to produce a proper weighting of the DNA evidence. As noted earlier, previous research suggests that jurors underweight statistical evidence, including DNA evidence. This would seem to bolster the argument that evidence should be presented in s/p rather than m/f form because the former is associated with greater perceived weight for the DNA evidence.

On the other hand, studies also show that people reason better with statistical data that are presented as frequencies rather than as probabilities.⁷⁰ People are not only more likely to reason in accordance with the logic of probability theory with frequencies than with probabilities, but a frequency presentation decreases the risk that people will commit various statistical fallacies.⁷¹

One reasonable solution is to employ frequency presentations in which efforts are made to discourage people from devaluing the DNA evidence on grounds of nonuniqueness. Alternatively, decision makers could be informed that there are different ways of presenting the same statistical information. These proposed solutions deserve empirical study.

C. Strategic Considerations for Attorneys

might misinterpret a likelihood ratio as a statements in favor of the odds that the suspect is guilty).

⁷⁰ Lea Cosmides and John Tooby, *Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions From the Literature on Judgment Under Uncertainty*, 58 COGNITION 1 (1996); Gerd Gigerenzer and Ulrich Hoffrage, *How to Improve Bayesian Reasoning Without Instructions: Frequency Formats*, 102 PSYCH. REV. 684 (1995); C. Harries and Nigel Harvey, *Are Absolute Frequencies, Relative Frequencies, or Both Effective in Reducing Cognitive Biases?* 13 J. BEHAV. DEC. MAK. 431 (2000).

⁷¹ Koehler 1996, *supra* at 880 (showing that mock jurors are more likely to commit source probability errors with DNA statistics that are framed as probabilities than those framed as frequencies); Thompson and Schumann, *supra*, at 172-6 (showing that the prosecutor's fallacy was committed by 22% of mock jurors who heard the statistical evidence in probability form but only 4% of mock jurors who heard the evidence in frequency form); Cosmides &

At a broad level, the message of these experiments for trial attorneys is that they should think not only about the objective strength of their statistical evidence but also about the imagery that their evidence evokes. For more than twenty years social scientists have known that laypeople underweight statistical information because it is dull and lacking in vividness.⁷² The present studies suggest that statistical information can overcome its natural propensity to be dull and unremarkable when it is presented in a way that promotes more concrete imagery about other exemplars.

At a slightly narrower level, the strategic implications of these data for trial attorneys and experts seem clear: prosecutors, prosecution experts, and others who wish to persuade others using DNA match evidence should favor an s/p (single target, probability frame) perspective for the DNA statistics: "The probability that the suspect would match the blood drops if he were not their source is 0.01%." In contrast, defense attorneys and defense experts who wish to cast doubt on the DNA match evidence should favor an m/f (multi target / frequency frame) perspective for the DNA statistics: "1 in 10,000 people in Houston who are not the source would also match the blood drops." This is particularly true in DNA cases involving relatively large match probabilities (e.g., one in hundreds or thousands) as opposed to very small match probabilities (e.g., one in millions or billions).

1. Prosecution: Beyond Single Targets and Probability Frames

Tooby, *supra*, (showing that probability reasoning errors reduced from 56% to 5% when frequencies are used).

⁷² Nisbett and Ross, *supra* at 55-59.

a. *Fractional numerators*

The underlying theoretical point in this paper is that presentations of statistical evidence that call match exemplars to mind will cause the evidence to seem weaker than presentations that do not call such exemplars to mind. A logical extension is that presentations that suggest exactly zero match exemplars should induce perceptions of great evidence strength. In a recent study, Laura Macchi and I showed that the ordinarily defense-friendly m/f method of presenting DNA statistics could be turned into a prosecution-friendly method by blocking people's cognitive access to exemplars.⁷³ We accomplished this by offering mathematically equivalent variations of the DNA match frequency statistic to hundreds of mock jurors. Three versions of two different DNA match incidence rates were used. For a one in one thousand incidence rate, the statistics were presented as "1 out of 1,000," "0.1 out of 100," or "2 out of 2,000." For a one in one hundred thousand incidence rate, the statistics were presented as "1 out of 100,000," "0.1 out of 10,000," or "2 out of 200,000." The three formulations within each incidence rate were mathematically, but not psychologically, identical. We predicted that the fractional numerators (0.1) would make it harder for people to think about others who might also match. Indeed, the near-zero numerators might even suggest that no one else is likely to match ("0.1 out of 10,000 sounds a lot like no matches to me"). In contrast, we predicted that numerators larger than one would *invite* thoughts of others who might match ("*Two* out of one hundred thousand? Hmm, I wonder which of the two is the source)."

As expected, we found that jurors were more impressed by the DNA evidence when it was presented with a fractional numerator and less impressed when it was presented with a numerator larger than one. This effect was so pronounced that jurors were more persuaded by "0.1 out of 100" than they were by 1 out of 100,000 and by 2 out of 200,000 (p(source)=63%, 53% and 46% respectively). In other words, jurors regarded objectively weaker evidence (0.1 out of 100) to be more persuasive than objectively stronger evidence!

2. *Defense: Beyond Multi Targets and Frequency Frames*

a. Tiny DNA match statistics don't account for laboratory error

When DNA match probabilities are extremely small (e.g., one in millions or billions), the prosecution obviously has the upper hand. However, defense attorneys may point out that those incidence rates do not account for the possibility that the reported match is not, in fact, a true match. When the possibility of a false positive laboratory error -- as estimated by laboratory proficiency tests⁷⁴ -- is taken into account, the underlying probative value of the reported DNA match may be more like one in several hundred or several thousand.⁷⁵ If defense attorneys can persuade jurors to think about DNA evidence not merely in terms of what it indicates about the

⁷³ Koehler & Macchi, 1997 paper presentation, *supra*.

⁷⁴ NRC I at 94 ("Proficiency tests provide a measure of the false positive and false negative rates of a laboratory").

⁷⁵ Koehler, Chia and Lindsey, *supra* at 210. Occasionally courts appreciate the important role that the possibility of error plays in the interpretation of DNA matches. See e.g., *Watts v. State*, 733 So. 2d 214, 224 (Miss. 1999) ("the introduction of statistical evidence can be meaningless without any evidence of the testing laboratory's error rate"). But see Margaret A. Berger, *Laboratory Error Seen Through the Lens of Science and Policy*, 30 U.C. DAVIS L. REV. 1081, 1090 (arguing that error rate estimates based on laboratory proficiency test results are not scientific).

possibility of a coincidental match, but what it indicates about the more inclusive possibility of a coincidental match *or* a false positive error, they may be able to invoke exemplar imagery

b. Use the largest plausible reference population

The evidence presented here shows that attorneys are not at the mercy of the DNA statistic. For one thing, they can frame the statistic. They can also invoke different sized reference populations to encourage or discourage the production of match exemplars. For a small match statistic (e.g., 1 in 1,000,000), a prosecuting attorney who invokes a relatively small reference population (e.g., the city of 40,000 people in which the crime occurred) makes it difficult for jurors to think about others who might match. This may be true even if the defense attorney invokes an m/f frame. Jurors may realize that the chance of locating even a single coincidental match in the city is so small that the DNA match statistic constitutes overwhelming proof of identity.

However, the defense attorney may wish to invoke a much larger reference population (e.g., the county or state). This can be justified by arguing that smaller reference classes presume that there is no possibility whatsoever that the donor of the genetic evidence came from outside the local area. Defense attorneys might do well to strike a balance here between choosing a reference class that is at once large enough to evoke match exemplars yet small enough so as not to seem absurd. In most cases, the largest reference class – all people in the world – would probably fail this intuitive test of absurdity. After all, it would strain credulity to suggest that a rape that occurred in a small town in Indiana might have been committed by any

of the 500,000,000 males in China.⁷⁶

VIII. FINAL COMMENT

This paper began with the assertion about the well-known power of DNA analysis in the courtroom. I caution that this power is a theoretical one. In practice, DNA analyses produce statistics that may or may not persuade jurors about whether a particular person is the source of a particular sample. The studies offered here suggest that the persuasive value of DNA evidence may depend as much on an understanding of the psychology of numbers as it does the underlying science and the statistical expression of that science.

⁷⁶ A large reference class might also lack credibility if it failed to exclude others who obviously could not have committed the rape such as women and children.