THE RULE OF PROBABILITIES: A PRACTICAL APPROACH FOR APPLYING BAYES' RULE TO THE ANALYSIS OF DNA EVIDENCE

Ian Ayres* & Barry Nalebuff**

Bayes' rule is not being used to guide jury decisionmaking in the vast majority of criminal cases involving evidence of DNA testing. Instead of telling juries the "source probability"—the probability that the individual whose DNA matches was the source of the forensic evidence found at the crime scene—experts only present pieces of the puzzle. They provide the probability that a randomly selected innocent person would have a match or the expected number of innocent matches in the database. In some cases, the random match probability will be so low (one in a quadrillion) that the resulting source probability is practically one hundred percent. But, in other cases, with large database trawls and random match probability at one in a million, jurors will have no ability to convert the random match probability or the likelihood ratio based on the expected number of matches into relevant data that will help them address the question of guilt. This Article shows that a correct application of Bayes' rule should lead factfinders and litigants to focus on the size of two variables that influence the source probability: the probability that a nonsource in the DNA database would have an alibi, and the probability that the source of the DNA is included in the database. This Article suggests practical means of estimating these two variables and argues that, as a legal matter, these parameters as well as the Bayesian posterior source probability are admissible in court. In particular, focusing on the prior probability that the "database is guilty" (i.e., the probability that someone in the database is the source of the forensic evidence) is not just analytically and empirically tractable, but avoids the evidentiary limitations concerning a particular defendant's prior bad acts. Appropriate application of Bayes' rule, far from preempting the factfinding and adversarial process, can guide advocates to engage the important aspects of the evidence that are still likely to be open to contestation. Perhaps most important, appropriate application of Bayes' rule will al-

---

so allow jurors to reach verdicts via a coherent path that employs sound logic and reasoning.

AN APPRECIATION OF RICHARD CRASWELL

It is entirely fitting for an Article on the legal implications of information economics to honor Dick Craswell.[1] Large swaths of Dick's writings are careful workings-out of the ways that imperfect information can impact private be-

---

1. Ian and Dick first connected in 1987, when Dick delivered a paper at the Law and Society Conference. Ian was just finishing his first year of teaching, and he hadn't yet overcome the kind of anxiety that caused him to tremble at the prospect of asking a question. Not only was Dick open and gentle in his response, but soon after the conference Ian received from him a letter of encouragement. Ian was floored by his letter because Dick, unbidden, had also included comments on a draft article of Ian's. That e-mail began a long-distance relationship that was immensely helpful to a young scholar. Over the decades, Dick and Ian mailed and e-mailed comments on over a dozen drafts of each other's writings. Though they never taught at the same school in the same semester, Ian nonetheless felt like Dick was one of his closest colleagues. Their academic *havruta* not only predates e-mail, it predates WordPerfect and Word—as those first comments were composed on the XyWrite word-processing software. No one else comes close to taking the time to have written as many or as high-quality comments on Ian's drafts. Dick has been a role model not only on the typed page, but also in the way he has carried himself in seminars, both while presenting and in the audience. Dick speaks softly but carries a collaborative and constructive stick. His ready smile is emblematic of his warmth. Just bringing to mind memories of Dick in seminar or talking over coffee lifts one's spirits. Of course, even very good academics can be acerbic and annoying. As with the Tsar, one sometimes wants to bless and keep them . . . far away. Not so Dick Craswell. He is proof that you can be a serious, respected scholar, *and* be universally admired. Even in the boisterous milieu of the University of Chicago seminar, he was a calming presence. There is a balm in Gilead. Richard Craswell.

havior and constrain judicial decisionmaking.[2] Dick's analysis of how consumers might mistakenly update their prior beliefs after a corrective advertising order[3] is a close analogy to our claim that unguided juries are likely to mistakenly update in response to incomplete statistical DNA evidence.

INTRODUCTION

With the recent Supreme Court decision allowing the collection of DNA samples from any person arrested and detained for a serious offense,[4] it seems inevitable that the justice system will collect and use large DNA databases. Currently, DNA databases are widely used. As of April 2015, the Combined DNA Index System (CODIS) maintained by the Federal Bureau of Investigation (FBI) had more than "283,440 hits [and had assisted] in more than 270,211 investigations."[5] There is concern that as database size increases, so too will the rate of false positives, and thus innocent people will be convicted when their DNA matches evidence left at a crime scene.[6] This concern has led courts to a convoluted and misguided use of multiple lenses to evaluate DNA evidence.

In this Article, we argue that there is a single right answer for how to incorporate the use of DNA evidence. That answer is the application of Bayes' rule, a 250-year-old formula for updating a starting probability estimate for a

---

2. *See, e.g.,* Richard Craswell & John E. Calfee, *Deterrence and Uncertain Legal Standards*, 2 J.L. ECON. & ORG. 279 (1986); Richard Craswell, *Taking Information Seriously: Misrepresentation and Nondisclosure in Contract Law and Elsewhere*, 92 VA. L. REV. 565 (2006); *see also* Richard Craswell, *Against Fuller and Perdue*, 67 U. CHI. L. REV. 99 (2000).

3. Richard Craswell, *Interpreting Deceptive Advertising*, 65 B.U. L. REV. 657, 689 (1985); *see also* Howard Beales, Richard Craswell & Steven C. Salop, *The Efficient Regulation of Consumer Information*, 24 J.L. & ECON. 491, 532-33 (1981) (discussing how consumers may update in response to banning certain advertising claims).

4. Maryland v. King, 133 S. Ct. 1958, 1980 (2013) ("When officers make an arrest supported by probable cause to hold for a serious offense and they bring the suspect to the station to be detained in custody, taking and analyzing a cheek swab of the arrestee's DNA is, like fingerprinting and photographing, a legitimate police booking procedure that is reasonable under the Fourth Amendment."). Under Maryland law, a serious offense is the commission of or attempt to commit violence or burglary. *See id.* at 1967.

5. *CODIS-NDIS Statistics*, FED. BUREAU INVESTIGATION, http://www.fbi.gov/about -us/lab/biometric-analysis/codis/ndis-statistics (last visited June 8, 2015).

6. There is an additional concern of false positives occurring not by chance but by contamination or laboratory error. *See, e.g.,* F.H.R. VINCENT, REPORT: INQUIRY INTO THE CIRCUMSTANCES THAT LED TO THE CONVICTION OF MR FARAH ABDULKADIR JAMA 21-23, 48 (2010); Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence at Trial*, 34 JURIMETRICS 21, 23 (1993); William C. Thompson, *Subjective Interpretation, Laboratory Error and the Value of Forensic DNA Evidence: Three Case Studies*, 96 GENET-ICA 153 (1995); Julie Szego, *Wrongfully Accused*, SYDNEY MORNING HERALD (Mar. 29, 2014), http://www.smh.com.au/national/wrongfully-accused-20140324-35cga.html. There is also the possibility of an innocent defendant's DNA being left at the scene by chance. *See* Jonathan J. Koehler, *DNA Matches and Statistics: Important Questions, Surprising Answers*, 76 JUDICATURE 224 (1993) [hereinafter Koehler, *DNA Matches and Statistics*].

hypothesis given additional evidence.[7] Applying Bayes' rule, we argue that triers of fact evaluating DNA evidence should be presented with what we call the "source probability": the probability that a defendant whose DNA matches the DNA found at the crime scene was the true source of that evidence. As we discuss below, the source probability is not the same as the chance of a random DNA match and does not equal the probability of guilt; even if the defendant was the source of the forensic DNA, the defendant might not have committed the crime.[8]

Our primary contribution will be to show that the source probability may turn crucially on the size of two variables that have not been introduced (or relied upon by experts) in DNA matching cases: (i) the initial or prior probability that the source of the DNA is included in the database, and (ii) the relevant or adjusted size of the DNA database, a calculation that takes into account the demographic information known about the criminal and the probability that a nonsource in the DNA database would have an alibi.

Experts have shied away from helping jurors form baseline beliefs, which are more formally called prior probabilities, and from then helping them convert those priors into a conclusion. The problem is that, absent priors, it is not clear how to coherently employ the expert information. As we discuss in our analysis of *People v. Puckett*,[9] an expert might well conclude that certain evidence makes it 100 times more likely that the suspect was at the scene of the crime. But 100 times more likely than what? The starting point or prior for a suspect who is identified from a large database trawl might well be less than 1 in 1000. In that case, 100-to-1 evidence is not persuasive. If the suspect was related to the victim and had motive and opportunity, then 100-to-1 would be much more convincing.

We will argue that there are practical means of estimating the prior probabilities and the relevant database size and that, as a legal matter, these parameters as well as the final source probability are admissible. In particular, chang-

---

7. *See* Thomas Bayes, *An Essay Towards Solving a Problem in the Doctrine of Chances*, 53 PHIL. TRANSACTIONS 378-81 (1763). We discuss Bayes' rule in Part II and note 43 below. Propositions 2, 3, and 4 in that Part present the rule. For an introduction to Bayes' rule, see ERIC GOSSETT, DISCRETE MATHEMATICS WITH PROOF 316-22 (2d ed. 2009). For an application of Bayes' rule to trial evidence, see David L. Faigman & A.J. Baglioni, Jr., *Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence*, 12 LAW & HUM. BEHAV. 1, 2 & n.1 (1988). For a colorful history of the rule's uses, see SHARON BERTSCH MCGRAYNE, THE THEORY THAT WOULD NOT DIE: HOW BAYES' RULE CRACKED THE ENIGMA CODE, HUNTED DOWN RUSSIAN SUBMARINES, & EMERGED TRIUMPHANT FROM TWO CENTURIES OF CONTROVERSY (2011).

8. Moreover, while we will speak of this source probability as the probability that the defendant was the source of the forensic DNA, the matching evidence cannot distinguish between identical twins and thus can only speak to the probability with regard to the defendant or the defendant's close genetic relatives. *See* Brief of 14 Scholars of Forensic Evidence as Amici Curiae Supporting Respondent at 34, *King*, 133 S. Ct. 1958 (No. 12-207), 2013 WL 476046; Koehler, *DNA Matches and Statistics*, *supra* note 6, at 224; Natalie Ram, *Fortuity and Forensic Familial Identification*, 63 STAN. L. REV. 751, 757, 758 n.35 (2011).

9. *See infra* notes 70-87.

ing the focus from a question about the prior probability that the defendant was the source to the prior probability that the "database is guilty"—that is, the probability that someone in the database is the source of the forensic evidence—not only is analytically and empirically more tractable, but also avoids the evidentiary limitations concerning a particular defendant's prior bad acts.

In *People v. Johnson*, a California Court of Appeal panel, in reviewing different types of DNA statistics, emphasized that "'the database is not on trial. Only the defendant is.' Thus, the question of how probable it is that the *defendant*, not the database, is the source of the crime scene DNA remains relevant."[10] But to apply Bayes' rule, the probability that the database contains the source of the forensic DNA, assessed prior to any consideration of whether an individual in the database actually matches, becomes a crucial input in determining the (posterior) likelihood that a particular matching defendant is the source of the forensic DNA. Contrary to *Johnson*, assessing the prior probability that the database includes the source—colloquially, "the probability that the database is guilty"—provides at once a readier means of estimation and a stronger argument for admissibility.

At the end of the day, we will acquit or convict a defendant, not a database. The problem is that it is very hard to directly estimate a starting point or prior probability for the likelihood that a specific defendant committed a crime. For example, what is the chance that some "John Doe" committed a crime before we have any evidence about Mr. Doe? In contrast, it is more coherent to ask the chance that a class of individuals, for example, convicted felons, would include the perpetrator of a crime.[11] For example, if half of rapes are committed by convicted felons, then the starting point would be fifty percent, assuming that the database contains all convicted felons. If jurors are to properly understand the implications of finding a match from a large database trawl, the size and characteristics of that database are relevant information.

Some legal analysts have been dismayed by the ways in which evidence of a DNA match tends to eclipse any role for adversarial engagement—turning litigants into little more than potted plants.[12] But appropriate application of

---

10. No. C068950, 2014 WL 4724896, at *20 (Cal. Ct. App. Sept. 24, 2014) (citations omitted) (quoting People v. Nelson, 185 P.3d 49, 66 (Cal. 2008)).

11. If the defendant is a member of that group, then we can convert the group probability to an individual probability by dividing by the size of the group or the relevant subset of that group that matches any known facts of the case. While we could call this the prior probability for the individual, the heavy lifting has all been done at the group level, and we should not pretend otherwise.

12. A rich scholarly literature investigates jurors' tendencies to over- or underestimate the probative value of DNA statistical evidence. *See* Jonathan J. Koehler, *The Psychology of Numbers in the Courtroom: How to Make DNA-Match Statistics Seem Impressive or Insufficient*, 74 S. CAL. L. REV. 1275, 1277-80 (2001); Jonathan J. Koehler & Laura Macchi, *Thinking About Low-Probability Events: An Exemplar-Cuing Theory*, 15 PSYCHOL. SCI. 540 (2004); Jonathan J. Koehler, *When Are People Persuaded by DNA Match Statistics?*, 25 LAW & HUM. BEHAV. 493 (2001); Jason Schklar & Shari Seidman Diamond, *Juror Reactions to DNA Evidence: Errors and Expectancies*, 23 LAW & HUM. BEHAV. 159 (1999). This lit-

Bayes' rule, far from preempting the factfinding process and the adversarial process, can guide advocates to engage with the important aspects of the evidence that are still likely to be open to contestation. We will show how estimation of both the prior probability and relevant database size can be assessed under alternative assumptions that are appropriately open to literal and figurative cross-examination to assure the robustness of the bottom-line conclusion: the defendant was or was not the true source of the crime scene evidence.

For more than forty years, scholars have been debating the appropriate use of probabilities and Bayesian inference in the courtroom.[13] Among the criticisms leveled at Bayesian reasoning is that jurors will be unable to integrate probabilistic and nonprobabilistic evidence, that the occurrence or nonoccurrence of a past event does not admit to intermediate probabilities, and that Bayesian inference is incompatible with our presumption of innocence or proof beyond reasonable doubt values.[14] Instead of engaging on the question of whether probabilistic evidence can or should elicit valid inferences of defendant guilt, we instead focus on a practicable way to present powerful evidence of the posterior probability that the defendant was the source of forensic evidence.[15]

In Part I, we discuss a motivating example that illuminates three conflicting approaches to statistical inference. Part II then lays out our notation and model applying Bayes' rule and explains why the two variables that we emphasize need to be accounted for. Part III analyzes the comparative statics of our model—how the source probability is affected by changes in five underlying parameters. Parts IV and V apply our Bayesian model to the cases of *People v.*

---

erature is related to a broader literature investigating jurors' problems in evaluating statistical evidence. *See* Jane Goodman, *Jurors' Comprehension and Assessment of Probabilistic Evidence*, 16 AM. J. TRIAL ADVOC. 361 (1992); D.H. Kaye & Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence?*, 154 J. ROYAL STAT. SOC'Y: SERIES A 75 (1991); Brian C. Smith et al., *Jurors' Use of Probabilistic Evidence*, 20 LAW & HUM. BEHAV. 49 (1996); William C. Thompson, *Are Juries Competent to Evaluate Statistical Evidence?*, LAW & CONTEMP. PROBS., Autumn 1989, at 9; William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 LAW & HUM. BEHAV. 167 (1987).

13. *See, e.g.*, Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970); Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1329, 1332 n.2, 1372 (1971).

14. *See* Tribe, *supra* note 13, at 1334, 1372 & n.137; *see also id.* at 1344 (attributing the view that probability evidence has "no application" in determining "what did or did not happen" because "it did or it didn't" to the New York Court of Appeals' decision in *People v. Risley*, 108 N.E. 200 (N.Y. 1915)).

15. We also ignore how biased DNA databases can inject inequality into the criminal justice system. For an extreme example, imagine a jurisdiction that only allows the DNA of minorities into its database. Our proposed methods would still produce accurate estimates of the posterior source probability whenever a defendant matched the database. But because only minority defendants were susceptible to such matching, the discriminatory database collection criterion might provide an equality rationale for excluding the evidence from trial.

Collins[16] and *People v. Puckett*, respectively. Part V explains how the underlying parameters in our model can be empirically estimated. Part VI discusses whether our approach is compatible with current rules of evidence.

I. THE ISLAND OF EDEN

Imagine that a singular crime has been committed on the otherwise idyllic island of Eden.[17] This island has a population of 51,295, and no one has come or gone since the crime. Thus, we know with certainty that the criminal is one of these 51,295 individuals. Moreover, given the nature of the crime, it is not possible to rule out anyone on the island as potential suspects. But there is one clue: the criminal has left behind a trace of DNA at the scene. Based on the distribution of DNA, there is a one-in-a-million chance that a random individual in the Eden population would match the DNA.[18] The people of Eden are upset about this crime and they agree that each individual will be required to provide a DNA sample to be tested. The elders of Eden are able to collect samples from 51,294 individuals, all but Mr. Baker. Unfortunately, Mr. Baker was killed in a fishing accident. This accident occurred after the crime but prior to the decision to collect DNA. Mr. Baker was given the traditional burial at sea and therefore his corpse is not available, and there are no personal items from which a DNA sample could be retrieved. There is no evidence to suggest that the tragic accident was in any way related to the crime or the subsequent investigation.

After collecting all of the DNA samples, it turns out that there was one—precisely one—match, a Mr. Fisher. Being trained in probability and statistics, Mr. Fisher was able to provide the police with the following argument and analysis:

> The test is too imprecise to use. What if Mr. Baker is the guilty party and thus all 51,294 of the remaining islanders are innocent? Because so many tests are done, there is a greater than five percent chance[19] that there will be at least

---

16. 438 P.2d 33 (Cal. 1968) (en banc).

17. We are not the first to use the "crime on an island" problem to illustrate the principles of DNA evidence and probability. *See* DAVID J. BALDING, WEIGHT-OF-EVIDENCE FOR FORENSIC DNA PROFILES 7-20 (2005); David J. Balding & Peter Donnelly, *Inference in Forensic Identification*, 158 J. ROYAL STAT. SOC'Y: SERIES A 21 (1995).

18. The probability is often much lower, but in this example there is only a small fragment, and so the probability is limited to one in a million. We also assume here that there is never a false negative: if the guilty person's DNA is tested, it will always match.

19. To be precise, the probability that the database will contain at least one match if Mr. Baker was the guilty party is equal to:

$$1 - \left(\frac{999,999}{1,000,000}\right)^{51,294} = 5.00007\%$$

This 5% includes a 0.127% chance that there will be two or more DNA matches. In these cases we assume that the court also convicts one of the two matches. If Eden will only make a conviction when there is exactly one match, then if the population of Eden expands to 52,707, the chance of exactly one match is 5.000006%, assuming there is just one person missing from the dataset and that person is the guilty party.

one match just by dumb luck. In this case, I was the unlucky fellow. But whether it was me or someone else, when everyone is innocent the chance is too high that someone will match. To state this more precisely, if the Eden court employs the test that a person is guilty if there is a match, then when Mr. Baker is the guilty party, the court will nonetheless convict an innocent person more than five percent of the time.

Eden's prosecutor was not convinced. Her reply was as follows:

I want to be able to sleep at night. The question that would keep me up is the thought of convicting an innocent person. It is true that if Mr. Baker is the guilty party, then there would be a five percent chance of making a mistake. But the chance Mr. Baker is guilty (before doing any DNA testing) is only 1 in 51,295. For there to be a miscarriage of justice, two things have to go wrong: the rare 1-in-51,295 event that Mr. Baker was the culprit and the event that (at least) one of the 51,294 living folks is an unlucky DNA match. Thus the procedure is fine. Before any DNA evidence was collected, the chance that some innocent person would end up being a match was below one in a million![20]

The judge took a third, slightly different view:

The question we have in front of us is the following: We have a specific person accused of a crime. What is the chance that Mr. Fisher is guilty? Is it beyond a reasonable doubt or not? That is the standard in Eden.

The probability that Mr. Fisher is the guilty party given that his DNA is the only match among the 51,294 samples can be determined as follows: At this point, there are only two possible suspects, Mr. Fisher and Mr. Baker. All of the other people in the population have been ruled out by the lack of a DNA match. Prior to the test, both Mr. Baker and Mr. Fisher were equally likely to have been the criminal. While we do not have additional information regarding Mr. Baker, we do know that Mr. Fisher had a match. The chance of this is one hundred percent if he is guilty and one in a million if he is innocent. Thus the odds are 1,000,000:1 in favor of his being the guilty party. We multiply this by the original odds of 1:1, and so the conclusion is that the revised odds are 1,000,000:1. Converting this to a probability, the chance that Mr. Fisher is guilty is 99.9999%.

The judge could also have reached this result using Bayes' rule.[21]

---

20. The chance that an innocent person will be convicted is 0.05/51,295, which is less than 0.0001%.

21. What are the prior odds that the database contains the guilty party? Initially, the odds that someone other than Mr. Baker is the guilty party are 51,294:1. To update beliefs, the judge would compare the observed number of matches to the expected number of matches that arise by chance even when everyone is innocent. The expected number of matches if everyone is innocent is 51,294/1,000,000, or 0.051. (This is equal to the population of the island multiplied by the probability that a random person will be a DNA match.) But there was one match observed, which is roughly twenty times more than what was expected if all living islanders were innocent. This increases the odds that someone other than Mr. Baker is the guilty party by almost a factor of twenty:

$$1 : \frac{51,294}{1,000,000}$$

Multiplying the two odds together, the revised odds that someone other than Mr. Baker is the guilty party are:

The calculations by the prosecutor and the judge are almost looking at the opposite sides of the same coin. The prosecutor is worried about implementing a rule that could end up convicting an innocent person. But the prosecutor's calculation is made before knowing whether or not a match will be found. In the present case, a match is very likely to be found (since the guilty party is very likely to be someone other than Mr. Baker). The judge is looking at the facts as they are, not as they might have been.

To see the folly of Mr. Fisher's argument, imagine that Mr. Baker was not lost at sea and that his DNA was included with everyone else's. And in this new scenario, Mr. Fisher remains the only match. Now we know with certainty that Mr. Fisher is the guilty party. Every other possible suspect has a perfect alibi: his or her DNA does not match. It is still the case that there is a five per-cent chance of finding a match by chance if everyone is innocent. But we know that not everyone is innocent: a crime was committed. If we would not reject the procedure in the case where all 51,295 DNA samples are collected, should we decide to abandon ship if only one sample is missing? We can no longer be certain that we have the guilty party in our pool of suspects, but the likelihood that we do is over 99.998%.

As the reader might have suspected, the population size of Eden was not chosen at random. Had the population been above 51,294, a test with a one-in-a-million chance of a false positive would lead to more than a five percent chance that at least one person would match even when everyone is innocent. Thus, once the population or database is above this size, there is little statistical power in the test of seeing a DNA match. If one employs a 95% confidence test, then Mr. Fisher (and classical statistics) argues that one should not engage in a trawl of a database larger than 51,294 (given the one-in-a-million random match probability).[22]

Our view is quite different. Whether the trawl is appropriate or not depends not just on the number of people in the database, but also on the chance the da-tabase contains the guilty party. Thus, if the entire population is 51,295, then it is entirely appropriate to trawl a database of 51,295 because the guilty party is in the database. Similarly, it is entirely appropriate to trawl a database of

---

$$51,294 \times 1:1 \times \frac{51,294}{1,000,000}$$

or 1,000,000:1, the exact same as in the judge's calculation.

22. We named our Mr. Fisher in honor of Ronald Fisher (1890-1962), the father of classical statistics. Under the null hypothesis that all living people on the island are innocent, we will still get a match more than 5% of the time. Thus, a match is not enough to conclude that we have a true positive if we require a 95% confidence level. Defendants, attempting to suppress DNA evidence, have raised this argument. *See, e.g.*, People v. Nelson, 185 P.3d 49, 61-62 (Cal. 2008) ("[W]hen a suspect is found by a search of a large DNA database, the chance of a coincidental match is increased because 'a single genetic profile (from the crime scene evidence) is compared to the very large number of profiles in these databases.'").

1,000,000 if the entire population is 1,000,000.[23] What matters is the probability that the guilty party is in the database.

Returning to Eden, we do not have certainty, as the database is 51,294 out of 51,295, but it is close enough and would be for many numbers below 51,294 as well. That said, we should be explicit about a key assumption implicit in this statement: there is no special connection between the person who is missing and the crime.

The fact that the person is not in the database is not evidence of that person's innocence or guilt.[24] For example, if a remorseful Mr. Baker were to have committed suicide by drowning, then Mr. Baker's "fishing accident" wouldn't really have been an accident. His premature death would have been an indication that he was the guilty party.

Moving from Eden and back to reality, these three arguments illustrate three potentially different answers in response to three different questions: First, how likely are we to find a match if everyone is innocent? Second, how likely are we to find a single innocent match? And third, how likely is it that a single match is actually innocent? Our view is that courts should answer the third question, which is the judge's approach, although this is something courts have been reluctant to do, since it may require coming to an initial view about the defendant's chance of being guilty *prior* to hearing the evidence of the case.[25] In DNA cases where the suspect has been identified through other means, the prior can be thought of as the chance the defendant is guilty based on factors such as eyewitness accounts, motive, and opportunity.[26] But in cases where the suspect has been identified via a DNA trawl, the prior may be based on factors unrelated to the case at hand. What matters is not just the size of the

---

23. In this case we would expect to see two matches, one of whom is the true source and one of whom is a random match. If we only see one match, then that person must be the true match, as the database contains the true source. If there are two or more matches, then we know one of them is the true source, but the DNA test has no way to distinguish between them.

24. Normally, the database is not a random sample of the population, and an individual's knowledge of this may influence his or her prior beliefs. For example, depending on the crime in question, knowing that someone is in a database of convicted felons may suggest that this person is more or less likely than someone not in the database to have committed a particular crime.

25. For a discussion of the conditions under which courts are willing to admit so-called "base rate" evidence, see notes 159-66 below.

26. In *McDaniel v. Brown*, the rapist was arrested based on witness description, proximity, and connection with the victim. 558 U.S. 120, 122-23 (2010) (per curiam). After the arrest, there was confirmation via DNA testing. *Id.* at 123-24. The defendant might have been convicted even without the confirmation of DNA evidence. *Id.* at 124-25. Thus, the prior might have been 80% or higher. In this case, the state significantly mishandled the presentation of the DNA evidence. One issue was the chance that one of the defendant's brothers might also have been a DNA match. *See* David H. Kaye, *"False, but Highly Persuasive": How Wrong Were the Probability Estimates in* McDaniel v. Brown?, 108 MICH. L. REV. FIRST IMPRESSIONS 1, 4 (2009), http://repository.law.umich.edu/cgi/viewcontent.cgi?article=1059&context=mlr_fi. However, the prior chance for the brothers would have been exceedingly low, especially for a thirteen-year-old brother living 440 miles away. *Id.* at 6.

database but the probability it contains the source of the forensic DNA. This is a critical ingredient to answer the ultimate question of whether the person in front of the court is guilty or not.

II. BAYES FOR AN ERA OF BIG DATA

Court cases introducing DNA evidence have traditionally focused on three different numbers:

1. The random match probability: the probability that a randomly selected person will be a DNA match.
2. The database match probability: the probability that the database contains at least one match. In practice, courts use the expected number of innocent persons in the database who will be DNA matches.[27]
3. The Bayesian posterior probability: the probability that the defendant is the true source of the DNA found at the crime scene given that the defendant is a DNA match.[28]

As the courts have recognized, these different numbers answer different questions.[29] But our view is that only one question is relevant, namely, what is the Bayesian posterior probability? This question is answered by combining the

---

27. We note that while the courts generally refer to something called a database match probability, what they actually calculate is *not* a probability. Instead, it is the expected number of innocent matches when the true source is not in the database. This number is then misread as a probability (although it is an upper bound on the probability, as demonstrated in NAT'L RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE 134-35 (1996) [hereinafter NRC II]). For example, in *Puckett*, discussed later on, there were 338,711 names in the database, and the court used a one-in-1.1-million probability for a random match. *See infra* notes 76, 85 and accompanying text. The expected number of innocent matches is 0.31, just less than 1/3. *See infra* Part V. But the probability of observing precisely one match is not 1/3. It depends on the chance the guilty party is in the database. One could ask, if we know that the guilty party is not in the database, how likely is it that there will be precisely one person (who we know is innocent) that matches? The answer to that question is $338{,}710 \times 1/1{,}100{,}000 \times (1{,}099{,}999/1{,}100{,}000)^{338{,}710} = 23\%$. One might argue that 23% is close enough to 1/3 in that neither is large enough to obtain a conviction. Our point is twofold. First, the 1/3 number is not a probability, and it is not appropriate to refer to it as such. Indeed, if the expected number of innocent matches in the database were two, we would not say that the chance of a database match is 200% and thereby violate a fundamental tenet of probability that all probabilities must be at or below one. Second, the database probability is only an input to what we really care about and should not be used on its own to determine the relevancy of a DNA match.

28. People v. Nelson, 185 P.3d 49, 63 (Cal. 2008). A fourth approach mentioned in *Nelson* is the "confirmatory loci" approach, in which the researcher searches a database with one set of DNA loci and, if a tentative match is found, uses another set of loci to confirm the match. *See id.* at 62.

29. *See, e.g., id.* at 61-63 (describing and explaining the differences between four methods for calculating the significance of a DNA match); United States v. Jenkins, 887 A.2d 1013, 1018-20 (D.C. 2005). The random match probability is the chance someone will match given that he is innocent. This is *not* the same as the chance someone is innocent given that he matches. Mixing these two numbers is an example of the "prosecutor's fallacy." *See* Thompson & Schumann, *supra* note 12, at 170-71.

prior beliefs with the information value of a DNA match in connecting a sus-
pect to a crime.[30] The random match probability and the database probability
are both inputs to that equation, but they are not useful on their own.

The fact that random match probability and database match probability are
not useful on their own does not make these numbers irrelevant. As essential
inputs, they should be examined for their accuracy. If the inputs are inaccurate,
then the final result will also be inaccurate. Even when there is some uncertain-
ty about the inputs, we can still reach a conclusion if we can be confident that
the inputs are above some level. This Article describes how, in practical terms,
to convert the inputs into the number the trier of fact should care about.

The easiest and best way to clear up the confusion relating to how courts
should use evidence concerning DNA is to demonstrate the correct way to pro-
ceed. In the process, one can see the role played by the supporting statistics.

It will be of enormous help to introduce some notation:

S will stand for the result that the defendant is the source of DNA found at
the crime scene.[31] This is simply shorthand. Instead of saying "What is the
probability that the defendant is the source of DNA found at the crime sce-
ne?" we say "What is Probability(S)?" or even "P(S)."

$\sim S$ will stand for the event that the defendant is *not* the source of the
forensic evidence.

M' represents the number of matches found in the database. In most
cases considered, there was only one person with a DNA match
found. But a trawl of a large DNA database with tens of millions of
samples might well turn up more than a single match.[32] To be as gen-
eral as possible, we will not restrict M' to be one. After M' matches
have been identified, some of the people found may have an airtight
alibi. Thus we introduce the related variable M.

M is the number of matches found where the person does not have an
airtight alibi.

---

30. *See infra* text accompanying notes 39-42 (discussing how Bayes' rule combines
these two types of information).

31. As noted in *Nelson*, *S* does not mean that the defendant is guilty of the crime
committed. For example, in *Nelson*, after a DNA match was established, the defendant
claimed that his semen was present as a result of prior consensual sex the weekend before
the victim disappeared. *See* 185 P.3d at 53-54. Moreover, *S* cannot exclude identical twins.
All that the DNA evidence can establish is that the person or a close genetic relative was
present at the crime scene and left behind semen, blood, skin, or saliva.

32. Think of DNA as akin to a highly unusual last name. For example, a trawl of
Nalebuff is unlikely to have more than one match in a city or state, but if the database is ex-
panded to include everyone in the United States, then a search would find a handful of
Nalebuffs.

a is the probability that a random innocent person in the database has an airtight alibi. Examples include the following: the person was not born at the time of the crime, was the wrong gender, or was incarcerated at the time. The use of the alibi screen can be applied either before or after the screening is done. For example, the DNA lab could screen all samples and then afterward look to see if the match found has an airtight alibi. Alternatively, the database could be scrubbed in advance to eliminate all people in the set who have an airtight alibi. In practice, we expect that some of both approaches will be taken. The dataset might be reduced to Caucasian males prior to the screening, and then, postscreening, any matches might be checked for birthdates and incarceration records that provide an airtight alibi. Fortunately, for mathematical purposes, we need not care when the screening is done. It makes no difference to the results if the screening is done prior to DNA matching, afterward, or a combination of the two.

r is the "random match probability."[33] It represents the accuracy of the DNA match. We measure accuracy by the chance that a random innocent person will have a positive match. More formally, it is the probability of match conditional on a random individual not being the source of the forensic evidence. Thus, lower values of *r* arise with more accurate tests. Experts may argue about the value of *r* in a specific case. In *People v. Puckett*, where there were five and a half loci matches, *r* was in the vicinity of one in 1.1 million for U.S. Caucasians.[34] In another case, with thirteen loci matches, *r* could be as low as one in 4.3 quadrillion.[35] The specific value of *r* is adjusted for other factors that might be known about the assailant, such as race and gender. Note that we assume that if the DNA sample was actually present at the crime scene, then the probability of a match is 100%.[36]

D is the size of the database against which the DNA sample is run. If there is just a single suspect whose DNA is run, then $D = 1$. When the DNA sample is compared to a larger database, *D* will represent the size of that database. The expected number of innocent matches in the

---

33. *See* 3 DAVID H. KAYE ET AL., THE NEW WIGMORE: EXPERT EVIDENCE § 14.1 (2d ed. 2010).

34. *See infra* Part IV. A loci match is an indication that a short sequence of DNA, called a VNTR (variable number of tandem repeats), is repeated the same number of times in two DNA samples. The same base pair sequence might be repeated anywhere from 70 to 450 times at a specific site on the human genome. The half match arose because information was degraded on one of the two alleles at the D21 locus.

35. *See* People v. Johnson, 43 Cal. Rptr. 3d 587, 591 (Ct. App. 2006).

36. We thus assume there is no chance of false negatives.

database is rD.[37] In the literature, rD is sometimes referred to as the "database match probability."[38] In fact, rD is not a probability but is an upper bound on the probability of finding one or more matches.[39]

p is the chance that the source of the forensic evidence is in the database.[40] For example, if all we know about the database is that it contains a single person picked at random from the planet Earth, then p might be one in twelve billion.[41] (The guilty person might be dead and thus not one of the living seven billion people.) If, on the other hand, we know that the guilty party is a Caucasian male who was alive last week and the crime was committed in Los Angeles, then this can be used to help refine p quite considerably. For example, in such a case, if the DNA database consisted entirely of Asian women, then the probability that the guilty party is in the database would be zero, $p = 0$.

At this point, we are ready to present the main finding. Any probability result is conditional on what we know. We know that a crime occurred, that DNA was found at the crime scene, and that if the source of the DNA is in the database, then there will be at least one match. We want to know the probability that someone in the database is the source of forensic DNA, given that we observe a certain number of matches:

Posterior Source Probability = Probability $[S \mid M$ matches$]$ (1)

We start with the aggregate probability that someone in the database is the source. This probability will tell us a great deal about the posterior source probability with regard to every individual in the database. Because we are assuming no false negatives, the posterior source probability of all unmatched indi-

---

37. To be more precise, rD is the expected number of matches in the database in the event that the guilty party is not in the database. For large databases, this makes little difference in that this changes the size of the database by one.

38. *See* KAYE ET AL., *supra* note 33, § 14.4.2-.3.

39. *See* NRC II, *supra* note 27, at 135.

40. In Bayesian terms, p is the prior probability of S, or the prior source probability (not conditional on any testing) measured at the aggregate database level. For simplicity, it is often referred to as the "prior."

41. There were roughly 7.2 billion people alive in 2015. *U.S. and World Population Clock*, U.S. CENSUS BUREAU, http://www.census.gov/popclock (last visited June 8, 2015). But if the database also includes all people born in the previous century, then the total would be closer to 12 billion. There were approximately 11.7 billion births between 1900 and 1995. *See* Carl Haub, *How Many People Have Ever Lived on Earth?*, POPULATION REFERENCE BUREAU, http://www.prb.org/Publications/Articles/2002/HowManyPeopleHaveEverLivedon Earth.aspx (last updated Oct. 2011).

viduals is zero.[42] This means that if there are no matches in the database, then the updated or posterior database source probability and the updated source probability of every individual will be zero. If there is a single unalibied match, then the posterior database source probability will be entirely focused on that matching individual (and the remaining unmatching individuals in the database will have a zero source probability). Finally, if there are multiple M unalibied matches in the database, then the posterior source probability of these matching individuals will just be the database source probability divided by M.

To derive this posterior database source probability, we deploy Bayes' rule,[43] which tells us that the posterior odds of an event occurring will equal the prior odds (not conditioned on evidence of any matches) of someone in the database being the source multiplied by the relative likelihood ratio of observing M matches:

Posterior Odds
$$= \text{Prior Odds} \times \text{Relative Likelihood of Observing } M \text{ matches} \qquad (2)$$

An odds formula is the ratio of two probabilities, one based on the event being true and the other based on the event being false. If the chance that the true match is in the database is p, the odds are $p : (1 - p)$, which we will write as the ratio:

---

42. The rate of false negatives varies depending on the width of the DNA band, or "match window," used to match the suspect to the source. The wider the band, the fewer the number of false negatives. *See* D.H. Kaye, *The Relevance of "Matching" DNA: Is the Window Half Open or Half Shut?*, 85 J. CRIM. L. & CRIMINOLOGY 676, 680-81 (1995). In the case we consider, there was a DNA match in the database. While this does not rule out the possibility of a false negative, for that to have happened, we would have to have experienced *both* a false positive and a false negative at the same time. This is doubly unlikely, and thus ignoring this possibility has a de minimis effect on the result.

43. Bayes' rule is commonly expressed according to the following equation:

$$p(A \mid B) = \frac{p(A \,\&\, B)}{p(B)}$$

where the symbol $p(A \mid B)$ means the probability of A given that B has happened, $p(A \,\&\, B)$ is the probability of both A and B happening, and $p(B)$ is the probability of B, here the number of matches. The intuition for this formula follows from a Venn diagram:
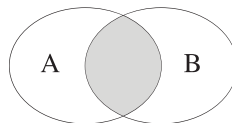
Once we know that B is true, then the only possible way for A to be true is that both A and B are true. Thus, to calculate the chance of A given B, we need only look at the fraction of times A and B are both true, limiting ourselves to the world in which B is true. An excellent introduction to the use of likelihood ratios and Bayes' rule in law is provided in Louis Kaplow, *Likelihood Ratio Tests and Legal Decision Rules*, 16 AM. L. & ECON. REV. 1 (2014).

$$\frac{p}{(1-p)} \tag{3}$$

Note that if the odds of a true positive are $A : B$, then the probability of a true positive is $A/(A + B)$. As expected, $p/(p + 1 - p) = p$, the prior probability of a database match.

Once we have new information, Bayes' rule provides the updated odds:

$$\frac{P(S \mid M)}{P(\sim S \mid M)} = \frac{P(S)}{P(\sim S)} \frac{P(M \mid S)}{P(M \mid \sim S)} \tag{4}$$

The odds start with our prior, namely, $p : (1 - p)$. But then we receive new information—there are M' matches of which M do not have alibis—and this allows us to update our odds.

The odds of observing M unalibied matches is the ratio of the probability this would happen when the source is in the database versus the probability this would happen when the source is not in the database (and the M unalibied matches occur by chance). Using the binomial distribution,[44] we show that the odds of observing M unalibied matches can be expressed in terms of M, D, r, and a[45]:

---

44. A classic example of the binomial distribution arises when a weighted coin is tossed N times. The binomial distribution provides the probability for each possible number of heads, from 0 to N. In our case, we use the binomial distribution for the probability that there will be M' random matches of individuals who were not the source, M without alibis, in a database of size D.

45. The odds are:

$$\binom{D-1}{M'-1}(1-r)^{D-M'}r^{M'-1}\binom{M'-1}{M-1}a^{M'-M}(1-a)^{M-1} : \binom{D}{M'}(1-r)^{D-M'}r^{M'}\binom{M'}{M}a^{M'-M}(1-a)^{M}$$

The left-hand side represents the binomial probability of observing $M - 1$ nonsource, unalibied matches from a database of size $D - 1$ (when the source is included in the database), while the right-hand side represents the probability of observing M nonsource, unalibied matches from a database of size D. Since odds are like a ratio, one can factor out common terms. In other words, 6:3 odds are the same as 2:1. Here, the odds ratio can be simplified to:

$$1 : \frac{D}{M'} r \frac{M'}{M}(1-a)$$

or

$$M : rD(1-a)$$

The odds are equal to the number of matches we actually observe that do not have alibis compared to the number we would expect to see by random chance. If it turns out that these two numbers perfectly coincide, then we do not update the prior probabilities. If the number of actual matches is large relative to the predicted number, then this makes it more likely that the match was a true match. Note that even if M is 1, this can still be a large number relative to the predicted number of innocent matches, $rD(1 - a)$, which could be 0.00000001.

We have been ignoring the possibility of a false negative. If the chance of this is z, then the revised odds would be:

$$M : rD(1-a) \tag{5}$$

This likelihood ratio can be restated simply as the ratio of the *actual* number of unalibied matches relative to the *expected* number of unalibied, non-source matches[46]:

$$M : E[M] \tag{6}$$

This likelihood ratio indicates how strong the new information is in terms of changing the prior opinion. For example, if the likelihood ratio in equation (6) is 10:1, then it is ten times more likely that the M matches observed are the result of the true match being in the database than all being there by luck. If our initial view was that it was twice as likely that the database did not contain the true match (prior odds are 1:2), then Bayes' rule tells us (via equation (4)) that putting these together means the new odds are 5:1 in favor of the database containing the true match.

Bayes' rule says that to derive the updated, posterior odds of the source being in the dataset, all we need to do is simply multiply the prior odds by the likelihood ratio of equation (6). Thus, by using Bayes' rule, the initial odds become the updated odds of:

$$\frac{p}{1-p}\frac{M}{rD(1-a)} = \frac{p}{1-p}\frac{M}{E[M]}$$

These posterior odds can be converted into the database source probability. If the odds are $A : B$, then the associated probability is $A/(A+B)$. In the present case:

Database Source Probability $[S \mid M] =$

$$\frac{pM}{pM+(1-p)rD(1-a)} = \frac{p}{p+(1-p)\dfrac{E[M]}{M}} \tag{7}$$

$$M + M\frac{arz}{1-r-z} : rD(1-a)$$

Note that this new expression will be nearly identical to the original odds. The new term is multiplied by a factor of rz, reflecting the fact that we have both a false positive and a false negative. When r and z are each small, the product will be vanishingly small and hence may be ignored.

46. The expected number of total (alibied and unalibied) innocent matches is rD, and the expected number of unalibied matches is merely this quantum multiplied by $(1-a)$.

This database source probability is the posterior probability conditional on ob-serving M unalibied matches that the source of the DNA is in the database (i.e., one of the M matching persons). The probability is a function of five underly-ing variables: M, r, D, a, and p.

A central takeaway of our analysis will be that courts and litigants should focus on the appropriate estimates for each variable. While the number of matches (M) and the random match probability (r) have standard means of es-timate, we will argue that the remaining three variables—the size of the data-base (D), the probability that a matching nonsource will have an alibi (a), and particularly the prior probability that the database includes the source (p)—have not been adequately theorized or empirically established in many court contexts.

The intermediate expression of equation (7) in terms of all five underlying parameters may look intimidating, but the expression on the far right has a more intuitive explanation. The posterior probability that the source was in the database is influenced by the ratio ($E[M]/M$) of how many innocent, unalibied matches are expected in the database ($E[M] = rD(1 - a)$) relative to how many unalibied matches are actually observed in the database (M). To the extent that the predicted number of unalibied, innocent matches is small relative to the ac-tual number of matches, we conclude that the match is likely to come from the true source. If one expected three innocent, unalibied matches and only ob-served one match, then the probability that the source is in the dataset will be lower than if one expected 0.01 innocent, unalibied matches but found one match.

Before looking at the full generality of this formula, let us begin with some simple cases. If $p = 1$, then equation (7) implies the database source probability is also equal to 1 as long as at least one person in the database is an unalibied match. This makes perfect sense. The case of $p = 1$ means that the prior proba-bility of the database including the DNA source is 100%, and as long as DNA testing uncovers at least one unalibied match, the posterior source probability remains at 100%.[47]

Another simple case arises when the expected number of unalibied, inno-cent matches ($E[M]$) is the same as the number of unalibied matches that are observed (M). In that case, the posterior database source probability remains unchanged by conditioning on the DNA evidence:

$$\text{Posterior Database Source Probability} = \text{Prior Source Probability} = p$$

We started out with a p chance that the source was in the database, and we end up in the very same spot. We expected (given our prior) to find a certain

---

47. If the database testing uncovers no unalibied matches, then equation (8) produces an undefined probability. This result also makes sense, as there is a contradiction between the assumed prior certainty that there is a match to be found ($p = 1$) and the exclusion of eve-ry individual in the database ($M = 0$).

number of innocent matches, and that is precisely the number we have found. If we had found more matches than we had expected, that would have increased the chance our original pool included the DNA source. If we had found fewer matches than expected, that would have decreased the posterior likelihood that our original pool included the DNA source. But if the observed number of matches is just equal to what we had predicted would arise by chance, then it is as if we had received no new information, and so our original (prior) estimate remains unchanged.

In many cases, when there is only one match ($M = 1$), the database source probability can be thought of as a function of just two variables, the prior probability that the source was in the database (p) and the expected number of innocent, unalibied matches ($E[M]$). If there is a 50% chance that the DNA source is in the database, and we estimate that the expected number of innocent, unalibied matches is 0.01, then the posterior probability that the source is in the database becomes more than 99%.[48] Intuitively, when Bayesians expect to find only 0.01 unalibied, innocent (nonsource) matches in the dataset, but then find one match, they update their prior belief, increasing the database source likelihood from 50% to 99%.

When the prior probability is (or is deemed as a matter of law to be) a very small fraction, equation (7) suggests that the odds of seeing M matches must be in the millions- or even billions-to-one range to have a good deal of confidence that the true match is in the database. For example, if a 100,000-sample database[49] were a random selection of the U.S. population, then p would be on the order of 0.03% (100,000/300,000,000), and the requisite odds of observing one innocent, unalibied match would need to be at least 300,000 to 1 to produce a posterior database source probability of 99%. And to get odds of 300,000:1 with a database of 100,000, one needs a random match probability of less than one in 300,000 × 100,000, which equals 1 in 30 billion.

Of course, the probability that the source is in the database is different from the probability that a particular individual is the source of the forensic data. Equation (8) provides the final step to convert to the probability that a particular person who matches is the source of the forensic evidence:

---

48. Database Source Probability $[S \mid M] =$

$$\frac{p}{p + (1-p)\frac{E[M]}{M}} = \frac{0.5}{0.5 + 0.5\frac{0.01}{1}} = 99.099\%$$

49. As of April 2015, the size of state databases ranged from 2,394,532 (California) down to 9693 (New Hampshire) profiles; the National DNA Index (NDIS), which contains DNA profiles contributed by federal, state, and local law enforcement agencies, contained 14,336,161 profiles. *See CODIS-NDIS Statistics, supra* note 5; *see also Frequently Asked Questions (FAQs) on the CODIS Program and the National DNA Index System*, FED. BUREAU INVESTIGATION, http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and -ndis-fact-sheet (last visited June 8, 2015).

$$\text{Source Probability } [S \mid M] = \left[\frac{1}{M}\right] \text{Database Source Probability} \qquad (8)$$

Equation (8) captures the simple point that when there are multiple matches, the DNA test evidence by itself cannot distinguish among the M people whose DNA matches the forensic DNA evidence. Thus, the source probability is divided equally between them, $1/M$ to each. If there is only one match, then the entire database source probability becomes the source probability for that matching individual. Thus, if we conclude there is greater than a 99% posterior probability that the source is in the dataset, and there is only one match, then there is a greater than 99% probability that the matching person was the source because all the other members of the database have been excluded.

III. COMPARATIVE STATICS

This Part explores how the source probability of equation (8) changes as we change the four underlying variables (a, r, D, and p) while holding M constant.[50] We also speculate on how these variables are likely to change over time. To begin, we rewrite the source probability formula, setting M to 1:

$$\text{Source Probability } [S \mid M = 1] = \frac{p}{p + (1-p)rD(1-a)} \qquad (9)$$

We use this formula to show how the source probability changes with respect to r, a, and p[51]:

1. Increasing the random match probability, r, while holding everything else equal decreases our confidence that the matching individual was the source of the forensic DNA. This makes intuitive sense. The larger the

---

50. It is also possible to estimate the impact of increasing the number of unalibied matches. Combining equations (7) and (8) leads to:

$$\text{Source Probability } [S \mid M] = \frac{p}{Mp + (1-p)rD(1-a)}$$

An increase in M while holding the other variables constant increases the denominator and thus lowers the probability that any of the individual matches is the true source. However, the aggregate probability that the set of people who match contains the true source goes up as the matching set of individuals increases.

51. The effect of a change in r and a is easily signed because, given that the numerator and denominator of equation (9) are positive, increases to the denominator necessarily reduce the source probability. To sign the effect of a change in p, we divide both the numerator and the denominator by p to get:

$$\text{Source Probability } [S \mid M = 1] = \frac{1}{1 + \dfrac{1-p}{p}rD(1-a)}$$

This makes it clear that as p increases, the denominator falls, and so the source probability increases.

chance that an innocent person will match, the lower the likelihood that a matching person will be the source of the forensic evidence.

2. If the chance of a randomly selected person in the database having an alibi increases, then, ceteris paribus, the chance of the matching person being the DNA source increases as well. As the alibi probability (a) increases, the chance of an unalibied, innocent match decreases and therefore the likelihood of an unalibied match being the source increases.

3. If all that changes is an increase in our prior, p, that the database includes the source of the forensic DNA, then once again it follows directly from the formula that the chance the source of the match is true increases. As intuition would suggest, a higher initial belief that the database includes the forensic source tends to increase our posterior belief that matching unalibied individuals are the true source.

Technology combined with law is likely to increase the probability that innocents will have alibis. Nonsources increasingly leave technological footprints of their whereabouts at particular times (for example, via iPhones and surveillance cameras).[52] This likely increase in the parameter a will consequently also tend to increase the database source probability.

The implication of a match from a larger database has been controversial.[53] The seemingly simple question is whether matches from a larger database produce a higher or lower source probability.[54] While this question can be easily answered if we apply our earlier assumption that all else is held constant, the problem is that increasing the size of the database also increases the probability that the source is part of the database. There are two opposing effects: a larger

---

52. *See, e.g.*, Jack M. Balkin, *The Constitution in the National Surveillance State*, 93 Minn. L. Rev. 1, 2 (2008) (discussing surveillance cameras); M. Wesley Clark, *Cell Phones as Tracking Devices*, 41 Val. U. L. Rev. 1413 (2007).

53. The National Research Council's 1996 report (NRC II) recommended multiplying the random match probability by the size of the database. NRC II, *supra* note 27, at 40.

This recommendation and Anders Stockmarr's proposal for a similar "sliding correction factor" to adjust for the size of the database, *see* Anders Stockmarr, *Likelihood Ratios for Evaluating DNA Evidence when the Suspect Is Found Through a Database Search*, 55 Biometrics 671, 674 (1999), led to an outpouring of responses and letters to the editor. *See* B. Devlin, Letter to the Editor, *The Evidentiary Value of a DNA Database Search*, 56 Biometrics 1276 (2000) (supporting Stockmarr's proposal); Peter Donnelly & Richard D. Friedman, *DNA Database Searches and the Legal Consumption of Scientific Evidence*, 97 Mich. L. Rev. 931, 950-52, 953 & n.65, 954-58 (1999) (criticizing NRC II and Stockmarr's proposal); I.W. Evett et al., Letter to the Editor, 56 Biometrics 1274 (2000) (criticizing Stockmarr's proposal); Ronald Meester & Marjan Sjerps, *The Evidential Value in the DNA Database Search Controversy and the Two-Stain Problem*, 59 Biometrics 727, 730 (2003) (concluding that "the database controversy [was] a *false* controversy"). For an overview of the debate, see David J. Balding, *The DNA Database Search Controversy*, 58 Biometrics 241 (2002).

54. The probability will strictly increase except in the extreme case where all of the additions are irrelevant.

database is more likely to contain the guilty party, but it is also more likely to lead to a match of an innocent person.[55]

Which of these two effects is more important depends on whether the new people added to the database were as likely as the original people to be the source of a match. The easy case is one in which everyone in the database has the same prior chance of being the source match, so that p is proportional to D. In that case, we show that the effect of the increase in p is (slightly) bigger than the reduction due to D, and thus we can be (slightly) more confident when a suspect is identified from a larger database.

A. *Trawling a Larger Database*

The question that often arises with a database trawl is how to adjust for the size of the database.[56] Is it more convincing to find a single match when trawling a database of size 1000 than when trawling a database of size 100,000?

To answer this question, we first assume that the two databases are each comprised of individuals who, from a Bayesian perspective, have identical prior likelihoods of being the source of the forensic DNA. In other words, the larger database has the same average quality as the smaller one in terms of finding matches. For example, two equal-sized cities might both have a database of convicted felons, but one city has been collecting data longer than the other.[57]

In this case, a match is more convincing from a database of 100,000 than from a database of 1000, but typically only by a small amount.[58] As it turns out, there are two forces that almost exactly cancel each other out.

On the one hand, when a single match is found in a database of size 1000, it is 100 times more informative than when it is found in a database of size 100,000. That is because the 100,000-size database should have 100 times as many random matches as a database of size 1000. *Thus, the odds are 100 times more in favor of guilt when we observe a match from a database of size 1000 than 100,000.*

---

55. A larger D generally leads to a higher p. But, holding p constant, increasing D increases the denominator and thus tends to reduce the source probability because it increases the likelihood that the match would be an unalibied, innocent person.

56. *See supra* note 53.

57. In this case, the two databases are not quite the same because the age distribution of people in the databases might be different. A larger database that adds in suspects who have a higher likelihood of being the source (because they are independently known to have been at the crime scene) or have ironclad alibis (e.g., because they were not born at the time of the crime) would have very different effects on the prior likelihood that the database includes the source.

58. *See* David J. Balding & Peter Donnelly, *Evaluating DNA Profile Evidence when the Suspect Is Identified Through a Database Search*, 41 J. FORENSIC SCI. 603, 604 (1996) [hereinafter Balding & Donnelly, *Evaluating DNA Profile Evidence*]; David J. Balding & Peter Donnelly, *How Convincing Is DNA Evidence?*, 368 NATURE 285, 286 (1994) [hereinafter Balding & Donnelly, *How Convincing Is DNA Evidence?*].

On the other hand, our prior belief goes almost exactly in the other direction. If we have 100,000 persons in the database, *the probability that the true match is in the database is 100 times greater than if we only have 1000 persons.*

Mathematically, one can show that the two effects don't perfectly cancel out: the prior *probability* is proportional to the size of the database, while the odds ratio is inversely proportional to the size of the database.[59]

If everyone in the database has the same prior chance of being the source, then $p = \alpha D$, and substituting this into equation (9) yields:

Source Probability $[S \mid M = 1]$

$$= \frac{\alpha D}{\alpha D + (1 - \alpha D)rD(1 - a)}$$

$$= \frac{\alpha}{\alpha + (1 - \alpha D)r(1 - a)} \qquad (10)$$

Note that the D terms all factor out, except for the $(1 - \alpha D)$ in the denominator. As D gets bigger, the denominator falls, and this means the final probability that the match is the source goes up. This proves the result.

However, the effect is very small until αD is close to 1. For example, if $\alpha = 1/T$ so that individuals are random draws of the population, and T is the population of California (37 million), then the right-hand side of equation (10) is only reduced by one percent in going from a database of size 1 to 370,000. In contrast, if the database includes all 37 million people, then the right-hand side goes to zero, and so the probability that the database match is the source goes to one. Provided that the database is a small fraction of the population and the database grows with constant "quality," then it is just slightly more convincing to

---

59. To show how the mathematics works, assume that each person in the database has the same chance of being the source. In that case, $p = \alpha D$. Let T represent the total population. If $\alpha > 1/T$, as might arise if this is a database of convicted felons, then the individuals in the database are more likely to be the source than a random person in the population. Similarly, if $\alpha < 1$, then the individuals in question are less likely than average to be the source. As claimed, the source probability is proportional to D. We can also convert this to odds. The odds ratio that the source is in the database—that the database is guilty, as we are calling it—is:

$$\alpha D : 1 - \alpha D$$

The updating to the odds if we get a single DNA match from a database of size D is:

$$1 : rD(1 - a)$$

where $rD(1 - a)$ is, as earlier, the expected number of nonsource, unalibied matches that would occur as a matter of chance. Observe that the expected number of random matches is proportional to D, so the odds are inversely proportional to D. Putting these two effects together, we get the odds the database is guilty if we see exactly one match:

$$\alpha D \times 1 : (1 - \alpha D) \times rD(1 - a)$$

We can factor out D from both sides of the ratio to get $\alpha : (1 - \alpha D)r(1 - a)$. Converting this back to a probability leads to equation (10).

find a single match in a large database. As the database becomes a large fraction of the total population (as was the case in Eden), then a match becomes much more convincing.

At the other extreme, consider the case where we add someone to the database who we know to be innocent but the courts do not. In this case, the chance that the source match is in the database is unchanged at p, but the expected number of innocent matches has gone up by r. Now the result of a trawl from a larger database is less convincing:

$$\text{Source Probability } [S \mid M = 1] = \frac{p}{p + (1-p)r(D+1)(1-a)} \tag{11}$$

An increase in D to $D + 1$, holding p constant, increases the denominator and reduces the posterior source probability.

Therefore, whether a match from a larger database is more or less convincing depends on whether the increased odds that the larger database contains the true source are more than enough to offset the expected number of innocent matches and thus the reduced odds that a match is the true source. The problem with increasing the size of the database is that a larger database will often be of lower quality when it comes to potential suspects. For example, adding random individuals to a database of convicted felons is likely to increase D more than it increases p. Of course, if enough people were added to the database so that it eventually became all-inclusive, then p would converge to 1 and the match would be perfect evidence.

B. *Tom Versus Tom, Dick, and Harry*

David Balding and Peter Donnelly make an argument that it is always better to observe a single match from a large database than a small one.[60] The intuition they present is as follows: Say that you tell me that Tom is a match. There is some chance that Tom is the true source. But it might also be Dick or Harry. Thus, consider the new result if we expand the database from just Tom to Tom plus Dick plus Harry. If the lack of a DNA match rules out Dick or Harry, then there are fewer other people who might be the true source. In this case, this evidence now points more toward Tom.

This argument is one hundred percent correct, but it doesn't fully answer the question of whether a match from a larger database is more or less convincing than a match from a smaller one.

The Balding and Donnelly result takes the view that the suspect (Tom) who has matched has some prior chance of being the true source and that prior prob-

---

60. "In summary, the strength of the DNA evidence, as measured by the likelihood ratio, is greater when the evidence results from a database search than in the probable cause setting." Balding & Donnelly, *Evaluating DNA Profile Evidence*, *supra* note 58, at 604.

ability does *not* depend on the size of the database. This seems like a fine assumption when Tom has been identified through other evidence, such as by an eyewitness. But if the only evidence against Tom is the DNA match from a database trawl, then the prior chance Tom is the true source may be influenced by the size of the database. Indeed, in the model for calculating priors that we present in Part V, the prior probability for Tom (and all others in the database) is the chance the *database* contains the true source divided by the relevant number of suspects in the database. And—this is critical—the chance the database contains the true match does not directly depend on the size of the database. (It depends on the nature of the criminal behavior at issue, which determines the lifetime chance a criminal will enter the database.) Thus, the prior probability for Tom is inversely proportional to the size of the database.

Here is another way of putting the difference. Start with the case where Tom is a felon and his DNA is a match. It is always true that the power of the match goes up if we test all other felons and none of their DNAs are a match. But the power of the match has to be applied to our starting point. What is the chance Tom will be a true match before the results of the test are known? This is Tom's prior. If Tom is related to the victim or has some motive, then the number of other felons might be irrelevant to the prior beliefs. But if we know almost nothing at all about the suspect (as in *People v. Puckett*, discussed in Part V below), then all relevant felons in the database (those with the right demographics and no alibi) are equally good suspects. But how good they are depends on the database size.

It is possible that a larger database has proportionally more good suspects, as might be the case if the larger database is more complete than the smaller one. Here the prior for any one suspect from a large database could be the same as the prior for any suspect from a smaller one.[61] But there are other situations where the prior beliefs for the database as a whole don't increase with size, and in that case the priors are less favorable when the suspect is a random member of a larger database.

The difference in answers is based on a difference in experiments. If we start with a suspect and a fixed prior belief about the suspect, then our beliefs become more certain to the extent that we can rule out a greater number of other individuals. More nonmatches is evidence that the match is correct. But if we don't start with a specific suspect and just ask how we feel about a match that comes from a large database trawl, then we cannot say that a match from a larger database increases our confidence that the match is truly from the source.

We want our approach to work for large databases as well as for very targeted searches. DNA testing of known suspects, which is also referred to as "confirmation" testing, presents a trawl of a very small database, often a single

---

61. In this case, the chance the database contains the true source grows linearly with the size of the database and a match from a larger database is more convincing; see the result following equation (10).

suspect.[62] We will have very different priors about the chance that the source is included in a test of five suspects than the chance that the source is included in a database trawl of 300,000 individuals, who as of yet are unconnected to the crime. We take up the question of how to implement the source probability equation for testing of a small database (that is, a case in which there is already a suspect whose DNA is being compared to the forensic sample) below in Part VI.B.

IV. APPLICATION TO *PEOPLE V. COLLINS*

Our analysis of DNA evidence can be usefully compared to the use of eyewitness evidence in the famous *People v. Collins* case. In 1968, the California Supreme Court overturned a jury verdict in which Malcolm Collins and his wife Janet were convicted of second-degree robbery.[63] The primary evidence against the couple was similar to a DNA match. The robbers were identified as an interracial couple, where the woman was blond with a ponytail, the man had a mustache and beard, and the couple drove a yellow car. Assuming that these features were all independent, the prosecutor's expert found a one-in-12-million chance that a couple would match this description. Many have commented on the implausibility of the independence assumption[64] and the challenge in finding the correct size of the relevant population.[65] The California Supreme Court also properly critiqued the lower court for not distinguishing between the random match probability (i.e., one in 12 million) and the probability of guilt. The Court calculated that the chance of finding two or more matches was 41.8%, which created reasonable doubt of the Collinses' guilt.

62. *See, e.g.*, David H. Kaye, *Trawling DNA Databases for Partial Matches: What Is the FBI Afraid of?*, 19 CORNELL J.L. & PUB. POL'Y 145, 155 (2009); William C. Thompson et al., *How the Probability of a False Positive Affects the Value of DNA Evidence*, 48 J. FORENSIC SCI. 1, 6 (2003); *see also* discussion *infra* Part VI.B. As DNA databases have expanded, there has been a marked rise in "cold hit" testing, in which a DNA sample is tested against a large database in order to find a match and therefore a potential suspect. *See* Andrea Roth, *Safety in Numbers? Deciding When DNA Alone Is Enough to Convict*, 85 N.Y.U. L. REV. 1130, 1140-41 (2010). Both "confirmation testing" and "cold hit" trawls are prevalent. *See, e.g., id.* at 1140 n.45 (providing evidence that the cumulative number of cold hits went from the double digits in 1996 to over 100,000 in 2010).

63. People v. Collins, 438 P.2d 33, 33 (Cal. 1968) (en banc). Note that only Malcolm Collins appealed the verdict.

64. *See, e.g.*, Finkelstein & Fairley, *supra* note 13, at 494-96; Tribe, *supra* note 13, at 1335-37.

65. The size of the relevant population should be chosen so that it contains the guilty couple. Since the crime was a purse snatching, it is unlikely the criminals felt a need to flee the state or even the city. Thus, the relevant population should be the number of couples in greater Los Angeles, where the crime was committed. *But a couple need not be married.* A couple could be any man and woman who know each other well enough to commit a robbery together, and this greatly increases the number of potential couples into the tens of millions. We show how this changes the estimates below.

We focus on two points not previously considered by the courts or commentators that flow from our analysis. The first point is that even the California Supreme Court did not pose the correct probability question. The court focused on the probability that there would be a second couple that matched given that the police had identified one couple that matched.[66] If there was another such couple, that would indicate that the Collinses might not be guilty. However, it is better to ask the right question directly: What is the probability that this particular couple present in court, which matches the characteristics and has no alibi, is the guilty party?[67]

In the case of a DNA match, we calculate the probability that the DNA-matching defendant is the true source of the forensic evidence. The translation to *Collins* is whether the characteristics-matching defendants (i.e., the defend-

---

66. A mistake made by the court (and subsequent commentators, *see, e.g.*, Tribe, *supra* note 13, at 1336 & n.23, 1337) is that it used the wrong formula for the question it tried to answer. Using our notation, the court asked, what is the chance that there is more than one couple that matches the description, given that we know that at least one couple does match? Applying Bayes' rule, the numerator is the chance that two or more couples match and the denominator is the chance that at least one couple matches:

$$\frac{1-(1-r)^T - Tr(1-r)^{T-1}}{1-(1-r)^T}$$

See Collins, 438 P.2d at 42-43. This would be the correct answer to the following question: Given a set of characteristics, how likely is it that two or more couples match by chance, given that at least one matches by chance in a population of size *T*? This is the wrong calculation for *Collins* because the police did not pick the characteristics first and look for matches second. Instead, the police picked the characteristics based on the description of the robbers. Thus, the likelihood that the guilty party matches is not *r*, but 100%. The calculation provided by the court is based on a model that all matches are random. But it is not random for the guilty party. That is the source of the mistake.

The correct chance that there will be two or more matches in the population as a whole is $1 - (1 - r)^{T-1}$. This is the chance that at least one innocent person will match all the characteristics by luck. Since we know there is a match for the guilty party, then there will be at least two matches in total if there are one or more innocent matches (or false positives). We emphasize this point because it is quite general. The fact that a crime has been committed means there is a guilty party that is a true positive in the population; this is relevant to all Bayesian calculations in criminal cases.

Our new formula is not quite the final answer. It is the probability of observing two or more matches *before* we have seen any matches. But the court did have a matching couple in front of them. If that couple is innocent, then the chance at least one more couple would match the characteristics is 100%, as the guilty couple will always match. If the couple in court is guilty, then the chance some other innocent couple will match is $1 - (1 - r)^T$, where *T* is here the number of couples not yet examined by the police. To calculate the correct probability that there would be another couple that matches, it would be necessary to average the two numbers, weighting the 100% result by the chance the couple is innocent and the $1 - (1 - r)^T$ probability by the chance the couple is guilty. This is much more complicated than simply calculating the chance the couple is guilty, especially as the guilt probability is an input to the above calculation.

67. Note that the court did not consider the chance that an innocent couple would have an alibi. In the case of the Collinses, they claimed to be visiting a friend, but the friend was vague as to the time of their visit. *See Collins*, 438 P.2d at 37.

ants whose characteristics match the eyewitness reports and who have no alibi) were the same couple that was seen at the crime. We calculate that probability below using the court's numbers as well as more reasonable estimates.

The calculation in *Collins* and any eyewitness match is different from a typical DNA trawl with respect to one very important feature. In DNA trawls, the police evaluate the entire database for matches. In contrast, with *Collins*, there is no practical opportunity to search the entire population of all couples in Los Angeles to see how many matches are to be found.

The probability of guilt in *Collins* thus depends on how far the police had gone in their search to find matching couples. If the police had searched the entire population of possible couples and found that the defendants were the only match, then we would know that the couple is guilty. The more the police have searched and only turned up one match, the more likely it is that the Collinses are guilty. The best case for the Collinses is that the police have not actually searched at all to see how many other couples fit the characteristics. We will proceed under this assumption to find a lower bound on the chance that the Collinses are guilty.

If there are T potential couples in the population and all of them are considered to be equally likely to be the guilty party, then our prior odds are $1 : T - 1$. (This is the equal-prior assumption.) The evidence value (likelihood ratio, in our earlier parlance) of a match when the defendant has no alibi is $1 : r(1 - a)$. Thus, the updated or posterior odds are $1 : r(T - 1)(1 - a)$. If $1/r$ and T are both large, this expression can be closely approximated by the odds $1 : rT(1 - a)$, which implies that the (posterior) probability of guilt is:

$$\frac{1}{1 + rT(1 - a)}$$

The numbers considered by the court were $r = 1/12$ million and $T = 12$ million. Assuming that the probability that an innocent person in the database has an airtight alibi is $a = 1/2$, we get $rT(1 - a) = 1/2$. We thus would expect to see 0.5 innocent matches and 1 guilty match, and so the chance of guilt would be $1/1.5 = 67\%$. The court erred in never calculating this posterior probability that the defendants were the same people seen at the crime scene.

Moreover, the numbers used by the court significantly underestimated the expected number of innocent couples that would match and so overestimated the probability of guilt. We can use demographic information to illustrate the problem. In the mid-1960s, there were roughly 100,000 black males in the relevant age group in greater Los Angeles.[68] This offers a better starting point for

---

68. The female robber was identified as being young. If we assume that her driver was in the seventeen-to-thirty-four age range, then roughly 1/3 of the 335,000 black males would be in the relevant age range. *See Historical Census Records of Ethnic Groups in Los Angeles County 1850 to 1960*, L.A. ALMANAC, http://www.laalmanac.com/population/po20.htm (last visited June 8, 2015); *Table 5: California—Race and Hispanic Origin for Selected Large*

the prior—namely, 1/100,000—given that the male robber was black. Suppose that we assume that (i) each black male knows 3 white women well enough to commit a robbery together, (ii) only 1 of 30 white women is blond with a pony-tail, (iii) 1/10 of cars are yellow, (iv) 1/10 of black men have a beard and mustache, and (v) 1/2 of the innocent population would not have an airtight alibi. Given these assumptions, the information value of a match falls to $r = 3 \times 1/30 \times 1/10 \times 1/10 \times 1/2 = 1/2000$. Even if these characteristics were all independent, we would still expect to find 100 couples that match the full description, 50 of whom do not have an alibi. Thus, the expected number of innocent couples that match would be closer to 50 than to 0.50. With an equal prior, the chance the Collinses are guilty would be $1/(1 + 100,000/2000) = 1/51$ or around 2%.

Our second point is that an equal prior is not the most reasonable starting point given all the information available to the factfinder. Even prior to the matching information, the Collinses seemed much more likely than an average couple to have committed the purse snatching. They had less than $12 to their names when they were married two weeks before the crime, and yet Malcolm Collins, who was unemployed, was able to pay $35 in traffic fines the day after the robbery (in which somewhere between $35 and $40 was stolen).[69] This shows the importance of using other information in forming the relevant prior or in updating the prior separate from the matching information.

One should start with the right question: What is the probability of guilt if a defendant matches the relevant characteristics and has no alibi? The answer depends on both the information from the match *and* the prior. The experts can do a far better job in presenting the relevant random match probabilities than was seen in *Collins*. With those probabilities in hand, calculating the chance of guilt is straightforward. In *Collins*, the correct calculation would have given the defense a strong argument that the probability of guilt was tiny, around 2%, unless one moves far away from a uniform or equal prior.

At the same time, experts helping jurors figure out an appropriate prior might still leave a jury with the confidence to convict. The Collinses were far from an average couple. The fact that they were dead broke just prior to the robbery and yet had unexplained spending right after the robbery should factor into the equation. Thus, enlightened application of Bayes' rule leaves ample room for advocates to engage on the most relevant issues regarding a defendant's innocence or guilt.

Cities and Other Places: Earliest Census to 1990, U.S. CENSUS BUREAU (July 13, 2005), http://www.census.gov/population/www/documentation/twps0076/CAtab.pdf.

    69. *Collins*, 438 P.2d at 34. The couple provided inconsistent answers as to where the $35 had come from. *Id.* at 35.

V. APPLICATION TO *PEOPLE V. PUCKETT*

On February 21, 2008, John Puckett was found guilty of first-degree murder for the 1972 death of Diana Sylvester.[70] Sylvester was a twenty-two-year-old nurse working at the University of California, San Francisco.[71] She was stabbed and strangled in her apartment.[72] The case remained unsolved for a number of years.[73] Then, in 2003, as part of a cold-case investigation, Bonnie Cheng performed a DNA analysis for the San Francisco Police Department (SFPD) using sperm samples obtained during the autopsy.[74] Due to the degradation of the samples, she was able to obtain useful data at five and a half loci.[75]

FIGURE 1

Defendant John Puckett, Right, in a San Francisco Courtroom with His
Attorney, Kwixuan H. Maloof

Photo: Robert Durell, *Los Angeles Times*

The DNA information was run through the California DNA database of convicted felons, which contained 338,711 profiles, and there was one match,

70. The facts of the case are taken from both the appellant's and the respondent's opening briefs and Chris Smith's article *DNA's Identity Crisis*. *See* Appellant's Opening Brief at 1-2, People v. Puckett, No. A121368 (Cal. Ct. App. Aug. 18, 2009); Respondent's Brief at 1-2, People v. Puckett, No. A121368 (Cal. Ct. App. Jan. 20, 2010); Chris Smith, *DNA's Identity Crisis*, S.F. MAG., Sept. 2008, at 66, 72, 82.

71. Appellant's Opening Brief, *supra* note 70, at 3; Respondent's Brief, *supra* note 70, at 1.

72. Appellant's Opening Brief, *supra* note 70, at 3; Respondent's Brief, *supra* note 70, at 2; Smith, *supra* note 70, at 72.

73. Respondent's Brief, *supra* note 70, at 2; *see* Appellant's Opening Brief, *supra* note 70, at 10.

74. Appellant's Opening Brief, *supra* note 70, at 10; *see* Respondent's Brief, *supra* note 70, at 3, 33.

75. Appellant's Opening Brief, *supra* note 70, at 12, 13 & n.10, 43-44, 66; Respondent's Brief, *supra* note 70, at 10 (quoting Ranajit Chakraborty & Jianye Ge, *Statistical Weight of a DNA Match in Cold-Hit Cases*, FORENSIC SCI. COMM. (July 2009), http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2009/undermicroscope/2009_07_micro01.htm); Smith, *supra* note 70, at 68, 74.

Puckett.[76] He was in the database because of a conviction for three separate crimes in 1977, in which he had abducted and raped two young women and sexually assaulted a third.[77]

The police located the then-seventy-two-year-old Puckett in a trailer park in Stockton, California.[78] They ascertained that he had been living in San Francisco in 1972 and subsequently arrested him.[79]

Puckett was convicted of first-degree murder and sentenced to life in prison.[80] The primary evidence in the case was his DNA match.[81] As reported by *San Francisco Magazine* contributing writer Chris Smith, who covered the trial, the other evidence was far from conclusive and might not even be considered incriminating.[82] Smith describes the ambiguous evidentiary record:

> [Lead homicide investigator] Toomey leans in and says, almost gently, "We have a DNA match, and it comes back to you." Puckett's reply: "I . . . I . . . don't remember this at all."
>
> So, is that a murderer's half-assed denial? Or the genuine protestation of a scared old man? It's like that all the way down the line: Puckett either matches the eyewitness description or doesn't. (Was he "medium-build with curly hair," as the description puts it, or "heavyset and balding," as photos from that time show him?)[83]

The jury also heard of Puckett's three prior rape and assault convictions. These prior convictions were admitted, notwithstanding the limitation on admitting prior bad acts, as the prosecutor argued that there was a common pattern in Puckett's crimes.[84]

---

76. *See* Appellant's Opening Brief, *supra* note 70, at 44.

77. *See id.* at 94-98; Respondent's Brief, *supra* note 70, at 5-6, 51; Smith, *supra* note 70, at 74. One of the victims persuaded Puckett not to proceed with intercourse due to an infection, at which point he forced her to orally copulate him. Appellant's Opening Brief, *supra* note 70, at 95.

78. *See* Smith, *supra* note 70, at 74.

79. *See* Appellant's Opening Brief, *supra* note 70, at 19; Smith, *supra* note 70, at 74.

80. *See* Appellant's Opening Brief, *supra* note 70, at 2; Respondent's Brief, *supra* note 70, at 1.

81. *See* Appellant's Opening Brief, *supra* note 70, at 49.

82. *See* Smith, *supra* note 70, at 78.

83. *See id.* at 78. At the end of the day, Puckett failed to present evidence of a convincing alibi, for example that he was incarcerated or lived in a different city or country. But that is not his burden (and it would be difficult for most people who had lived in San Francisco to explain in 2008 where they were on a particular night in 1972).

84. Section 1108 of the California Evidence Code permits the introduction of evidence of other "sexual . . . offenses" "[i]n a criminal action in which the defendant is accused of a sexual offense," provided that the evidence is not inadmissible under section 352's general exclusionary rule. CAL. EVID. CODE § 1108 (West 2015); *see also id.* § 352 (allowing a court to "exclude evidence if its probative value is substantially outweighed by the probability that its admission will (a) necessitate undue consumption of time or (b) create substantial danger of undue prejudice, of confusing the issues, or of misleading the jury"). Section 1108 is an exception to section 1101(a)'s prohibition on the introduction of "character" evidence to show propensity. *See id.* § 1101(a).

Thus, the jury was left with the following picture: There is a convicted triple rapist whose DNA matches that of the semen left at the crime. The chance a DNA match would happen by chance is below one in a million.[85] The jury was not told how the match was found, the size of the database, or any other information that would help them form a more accurate estimate of Puckett's likelihood of guilt. Absent a rock-solid alibi, it seems predictable that Puckett was found guilty, although it took the jury three days of deliberation to render the verdict.[86] That was the final ruling. An appeal to the California Court of Appeal ended when Puckett died on July 21, 2010.[87]

Our central claim is that the statistical evidence presented—chiefly the random match probability—was incomplete. Without putting the statistical evidence into context, the jury did not have enough information to reach a conclusion one way or the other. Some jurors, when told about statistics that were withheld, said that they would have reached a different verdict.[88]

In this Part, we show how our earlier analysis of the posterior source probability should have guided the way that the DNA evidence was presented to the jury. In our analysis, we begin by using the odds ratios and then show how

In *Puckett*, the court purported to analyze the evidence of Puckett's other sex crimes under both section 1108 and section 1101. Initially, the court admitted the evidence as propensity evidence under section 1108. Respondent's Brief, *supra* note 70, at 51. In reaching this conclusion, the court weighed the section 352 factors, including whether the other acts would be "unduly inflammatory" or would cause confusion among the jurors, the remoteness of the other acts, and the similarities between the other crimes and the crime charged. *Id.*

Upon motion from the defense, the trial court reanalyzed the other-crimes evidence under section 1101, but held that its ruling on the "1101 issues . . . remain[ed] the same," and that the evidence should be admitted to "show common scheme." *Id.* (internal quotation marks omitted). At trial, the court instructed the jury that it could consider the evidence "for the limited purposes of deciding whether or not the defendant had a plan or scheme to commit the sex offenses claimed in this case." *Id.* at 52 (internal quotation marks omitted).

The prosecution described the common scheme in Puckett's sex crimes as follows:

He told a number of his victims, before he raped them, that he wanted to "make love." This statement, at least in prosecutor Merin's eyes, is uncannily similar to what Sylvester's killer said—"Go away, we're making love"—as he chased away her landlady, the only one who actually saw him, before slamming the door in her face.

Smith, *supra* note 70, at 78.

On appeal, the defense argued that the other-crimes evidence was insufficiently similar to the crime charged and unduly prejudicial and that its admission constituted reversible error. *See* Appellant's Opening Brief, *supra* note 70, at 93-108.

85. Bonnie Cheng, the government's SFPD criminal lab technician, testified that the genetic profile would occur at random among unrelated individuals in about one in 1.1 million U.S. Caucasians. The defendant's expert, Laurence Mueller, criticized this calculation. *See* Respondent's Brief, *supra* note 70, at 4, 9, 33-34.

86. *See* Smith, *supra* note 70, at 82.

87. *See* Abatement of Appeal, People v. Puckett, No. A121368 (Cal. Ct. App. Oct. 6, 2010).

88. One of the jurors, Joe DeLuca, was told after the trial about the "one-in-three" chance of a random match. His response: "That would have changed everything." Smith, *supra* note 70, at 82 (internal quotation marks omitted). As will be clear, we think the one-in-three statistic for a random match is incorrect and thus might have changed things the wrong way. We don't think that providing bad data is the solution for providing no data.

these odds estimates can be converted into an estimate that the defendant was the source of the forensic DNA.

In the present case, there was strong circumstantial evidence that the source of the forensic evidence was guilty of rape and murder. Prosecutor David Merin made this connection in simple and vivid terms: "His DNA was found in her mouth."[89] Thus, in this case, source and guilt probabilities are the same. Puckett did not claim that he had consensual sex with Sylvester or that the DNA was that of a sibling.[90] Thus, for this case, we will use the source probability as the probability of guilt in our discussions below.

To foreshadow the results, we think the evidence correctly presented is still strongly in favor of Puckett's guilt. There are three important caveats to this conclusion. First, some of the analysis we use, especially the formation of prior beliefs, might not be admissible. We return to this important question below in Part VI, where we will argue that, notwithstanding appropriate limitations on introducing evidence of prior bad acts, courts should be able to hear expert testimony about the Bayesian posterior (and the underlying prior on which it is based). Second, we don't have access to the actual DNA database, and thus several of our calculations are estimates. We try to be conservative in our estimates. But the data do exist, and thus our calculations suggest a method that can be improved upon. The next Part will make explicit how this can be done more generally. Finally, we discuss the role of the first suspect in the case, Robert Baker, and how that might influence our view of Puckett's potential guilt.

Separate from our reflections on Puckett's guilt, this analysis will suggest a new approach to deal with hits reached from a large database trawl. As a starting point, we believe that the jury should not have been told only that the chance of a random U.S. Caucasian matching at the five and a half loci was one in 1.1 million. They should also have been told that the DNA match was found as a result of a database trawl of 338,711 felons.[91] In terms of our earlier notation, the jurors should have been presented with D in addition to r.

But that does not mean that the chance of a random match, the "random match probability," was one in three. First, $338,711/1,100,000 = 0.31$ is the expected number of random matches (what we call rD in our earlier analysis). It is *not* the probability of at least one match if everyone in the database is randomly selected from the population of innocent persons. The actual probability is:

89. Jason Felch & Maura Dolan, *When a Match Is Far from a Lock*, L.A. TIMES (May 4, 2008), http://articles.latimes.com/2008/may/04/local/me-dna4 (internal quotation marks omitted).

90. *See* Appellant's Opening Brief, *supra* note 70, at 23-26; Respondent's Brief, *supra* note 70, at 6-7.

91. *See* Appellant's Opening Brief, *supra* note 70, at 44; *see also* Respondent's Brief, *supra* note 70, at 14, 24.

$$1-\left(\frac{1,099,999}{1,100,000}\right)^{338,711} = 26.5\% \ [92]$$

While 26.5% is smaller than 33%, this is a minor point, and not one that changes any of our conclusions. But we want to start with the correct calculations.

The reason to change conclusions is that 26.5% grossly overstates the chance of a random match when excludable matches are removed from the database. Since an eyewitness identified the criminal as a male Caucasian, had a female or a non-Caucasian matched, we would then know this is an innocent match. Thus, only the male Caucasians in the database are relevant.

According to the State's brief in Puckett's appeal, 28.4% of inmates are Caucasian.[93] If this percentage was constant over time, then only 96,194 felons in the database are relevant. This lowers the chance of a random match to 8.4%. We should also eliminate women from the database. We don't have access to the data, but in 1990, 14% of felony defendants were female according to a report by the Bureau of Justice Statistics.[94]

Furthermore, we believe that the criminal was at least age eighteen at the time of the attack, which implies an age of at least fifty at the time of the DNA search. We don't know how many people in the database were older than fifty in 2004. This is a knowable fact, but we do not have access to this information. The State argued that, as of December 31, 2005, just over 5% of inmates were over fifty years old.[95] However, that is not a relevant statistic. We are not concerned with the chance that the guilty party is presently incarcerated. We are concerned with the age distribution of felons in the database, recognizing that DNA could have been entered some ten or twenty years prior. These data do suggest that the fraction over fifty should be at least 5% because the current figure is 5% and data entered from earlier periods would have had a chance to age up. For purposes of illustration, we will assume that the age distribution of Caucasians in the felon database mirrors the overall Caucasian population that is twenty-five and over.[96] From 2000 census data, 42.7% of the Caucasian male population that is at least twenty-five is age fifty or older.[97]

---

92. The probability that a single randomly selected innocent person will not match is 1,099,999/1,100,000. The probability that 338,711 randomly selected innocent persons will not match is this ratio raised to the power of 338,711. Therefore, the probability of one or more matches among 338,711 randomly selected innocent persons is this quantum subtracted from 100%.

93. *See* Respondent's Brief, *supra* note 70, at 15.

94. *See* BRIAN A. REAVES, FELONY DEFENDANTS IN LARGE URBAN COUNTIES, 2009— STATISTICAL TABLES 1 (2013), *available at* http://www.bjs.gov/content/pub/pdf/fdluc09.pdf.

95. *See* Respondent's Brief, *supra* note 70, at 14.

96. We do not expect the felon database to include infants and teens. To be conservative, we looked at the age distribution among those twenty-five and older. Although we have been hypothesizing about the characteristics of the DNA database, this is something that can

Assuming the database in Puckett's case was a representative sample of the overall population of felony defendants nationwide, taking those under fifty and women out of the database lowers the relevant database population to $96{,}194 \times 0.427 \times 0.86 = 35{,}325$. This estimate deflates the database size to those without alibis—as captured in equation (7) by $D(1 - a)$. This netting out of alibied members of the database reduces the chance of an innocent match to something below 3.2%.[98] In other words, the expected number of matches without an airtight alibi is only one-tenth as many as the defense claimed.

But, as argued above, we think it is not productive to focus on the chance of a random innocent match. Instead, we should look at the strength of the evidence that follows from a match. The likelihood ratio used to update the prior is the observed number of matches relative to the predicted number of innocent matches, $M : E[M]$. In this case, with 35,325 viable felons in the database, the predicted number of male matches who are Caucasian and older than fifty is $35{,}325/1{,}100{,}000 = 0.032$. Thus, it might seem that the likelihood of observing the match in favor of Puckett being the source compared to an innocent non-source match is 1:0.032 or 31:1.

But even this calculation significantly understates the likelihood of Puckett being the source of the forensic evidence. The calculations above did not consider other potential airtight alibis, such as proof that the person was in another state or was incarcerated at the time the crime was committed. Since this was a database of felons, many of them will have been convicted more than once. For purposes of illustration, we will assume that felons spend 33% of their adult life in prison.[99] We will also assume that a random innocent person in the database will have a 50% chance of finding some definitive piece of evidence that shows he was not in the relevant geographic area at the time.[100]

We understand that Puckett was entered into the database three times, once for each of his convictions. Thus we expect that the database will have multiple duplicates; for illustration we will assume that 25% of the names in the database are duplicates. That reduces the size of the relevant database to $0.67 \times 0.5 \times 0.75 \times 35{,}325 = 8875$ and the expected number of innocent matches to $E[M] = rD(1 - a) = 0.008$.

---

be known. With access to the database, we can figure out precisely how many of the 338,771 names are white males over fifty.

97. *See Table 2: Male Population by Age, Race and Hispanic or Latino Origin for the United States: 2000*, U.S. CENSUS BUREAU (Oct. 3, 2001), http://www.census.gov/population /cen2000/phc-t9/tab02.pdf.

98. $35{,}325/1{,}100{,}000 = 0.032$.

99. Eleven percent of felons have five or more convictions. REAVES, *supra* note 94, at 8. The average sentence for the most serious felony offense is approximately fifty-two months, *id.* at 30, and roughly sixty-two percent of felony defendants had prior felony arrest charges, *id.* at 11.

100. While we have made up this percentage for the purposes of illustration, we note that it can be empirically estimated. One could randomly sample database members and ascertain whether they have a strong alibi, for example because they were incarcerated or lived out of state.

Now the odds in favor of Puckett being the guilty party compared to a random match are 1:0.008 or 125:1.

This is much more convincing than the defense estimate of 3:1, but it is still incomplete. To estimate the posterior source probability, we need also to combine this evidence from the DNA match with the initial chance that the guilty party was in the database. Thus, along with information about D and r, jurors should also learn about p, the variable that represents the chance the database contains the guilty party.

We appreciate that it is controversial to consider the prior odds that a defendant is guilty or that a database pool contains the guilty party. But if we don't take that into account, then we are left with the following: based on the DNA match, the database as represented by Puckett is 125 times more likely to contain the guilty party than what we thought prior to finding a match. But 125 times more likely than what? Without a prior, that number exists in a vacuum and doesn't help us reach a conclusion.

To reach a conclusion, we can work backward: We first ask what are the minimal priors required to eliminate reasonable doubt. Even if the priors are imprecise, provided they are above this threshold, we can reach a conclusion.

We then proceed with a new approach for calculating priors, based on a simple model of criminal behavior that leads to predictions for the fraction of criminals that are caught and thus appear in a database.

A. *Minimal Priors*

Before any match is found, the initial odds that the database contains the guilty party were:

$$p:1-p$$

With the new data, the updated (or posterior Bayesian) odds become:

$$125 \times p:1 \times (1-p) \text{ or } 125p:1-p \tag{12}$$

Associated with these odds is a probability that Puckett is the guilty party. This is:

$$\text{Source Probability} = \frac{125p}{125p+1-p} \tag{13}$$

If we imagine that proof beyond reasonable doubt requires establishing a source probability at or above 99%, then we can work backward to derive a minimum prior that would produce that posterior probability range:

$$\frac{125p}{1+124p} \geq 0.99$$

That is:

$$2.24p \geq 0.99 \text{, or}$$
$$p \geq 44\%$$

Our threshold is 44%. If we believe there is a 44% or higher chance that the guilty party is in the 338,711-felon database, we can conclude that Puckett has a 99% or higher chance of being the person whose DNA was left at the crime scene.

Is 44% a reasonable number or not?

B. *A Model for Calculating Priors*

Certainly a random person on the street would not have a 44% chance of being the guilty party. But the 338,711 felons are not random, and they form a large group. Even so, those two features alone are not enough to guide a juror (if the juror were even allowed to know these facts) in determining whether 44% is reasonable for the prior chance the true source is in the database. We should not count on intuition or prejudice.[101] To answer that question, we need a framework to calculate the prior probability.

One approach to estimating the prior would be to compare the size of the database to the size of the population without alibis. This method, by shrinking the size of the relevant database from 338,711 to 8875, would greatly reduce the chance that the guilty party will be found in the data. But as we have shrunk the relevant database, we have also shrunk the relevant comparison set. Thus, instead of comparing 338,711 to the overall 19.9 million total population of California at the time, one can compare 8875 to the Caucasian male population over fifty without an alibi for December 22, 1972. Even with a reduced denominator, this approach to estimating the prior would produce an estimate well below 44%.

We don't think this is the correct way to proceed. It is not enough to reduce the relevant geography to the population of California or even San Francisco. The problem is in the assumption that convicted felons have the same probability of committing a crime as a random person living in California. We recog-

---

101. A prejudiced juror might conclude that because Puckett was a convicted rapist, his prior chance of being guilty could be close to 50%. The problem with this argument is that all the individuals in the database were convicted of some felony and thus could all be labeled bad actors, each with a 50% chance. But the combined prior probability for *all* 8875 felons in the database with appropriate demographics and no alibis must add up to something less than one.

nize that courts have been reluctant, as discussed elsewhere in Part IV, to admit evidence of prior bad acts as evidence of a defendant's heightened propensity of committing a crime. We think it will be less prejudicial to focus on the database rather than the specific defendant, and thus we consider the probability that the criminal exists in the database prior to any DNA testing. This is another way of saying what fraction of criminals eventually end up in the database. To calculate this probability, we build a model of how criminals enter the database.

We will employ a simple model, but one that is realistic enough to help us form a practical estimate. Our model starts with the criminals active in 1972, what we will call the criminal class of 1972. We know that one of these criminals was the guilty party. How likely is a member of this criminal group to have ended up in the database by 2004? That is precisely the number we are looking for.

We assume that criminals behave in the following manner. In year 1 (1972), each criminal in the class commits a crime. For simplicity, we can say there are 100 criminals in the group, and in so doing our estimates will be transformed into a percentage rather than a number.

Of the 100 criminals, fraction f are caught and $(1 - f)$ are not. Those who are caught are entered into the DNA database as convicted felons.[102] This leaves $100 \times (1 - f)$ criminals who are not in the database. Some fraction d of these unconvicted criminals die, are killed, retire, or move to a different jurisdiction. Those who do will never be entered into the database. The fraction $(1 - d)$ who don't die, move, or retire then go on to commit another crime. Again, fraction f are caught and $(1 - f)$ are not. Those caught are entered into the database, and those who have escaped conviction twice are still not in the database. After this second crime, again a fraction d of those who have escaped conviction die, retire, or move. This continues until the original population has entirely been convicted or exited from the population via the d route.

We don't have to specify if criminals commit crimes once a year or once a month or something else. For our purposes, we have a group of criminals from the year 1972, and we want to see where they have ended up by 2004. Even with a relatively long period between crimes (one year), nearly all of them will have been convicted or "retired" thirty-two years later. Those who "retired" prior to being convicted will never enter the database, while those who are convicted prior to retiring will be in the database. In terms of probability, it is a race between being caught and "retiring."

Our focus is on the $(1 - f)$ criminals who are not caught in their first crime.[103] We want to know how many of these unsolved crimes are committed

---

102. At present, all those arrested are entered into the database.

103. The perpetrator of Diana Sylvester's murder was not caught at the time. Thus, for the perpetrator to show up in the felon database, the person would have to have gotten away with a crime and then been caught in a subsequent crime. It is also possible that the perpetrator might have already been in the database for a crime committed prior to Sylvester's murder in 1972. While we ignore that possibility here, this would only increase the chance that

by someone who later on ends up in the database. Fraction $(1 - d)$ of them will commit another crime, at which point f will be caught. Fraction $(1 - d)^2$ will commit a third crime, and of the $(1 - f)$ not caught the second time, f will be caught on the third go-round. Solving an infinite series,[104] we can estimate the chance that one of these criminals from the class of 1972 will end up in the database:

$$p = (1-d)f + (1-d)^2(1-f)f + (1-d)^3(1-f)^2 f + \ldots$$
$$= \frac{(1-d)f}{1-(1-f)(1-d)} \tag{14}$$

Our condition on the prior probability p can be rewritten as a function of f and d:

$$\frac{(1-d)f}{1-(1-f)(1-d)} \geq 0.44 \tag{15}$$

$$0.56(1-d)f \geq 0.44(1-(1-d))$$

$$0.56f \geq \frac{0.44d}{1-d}$$

$$1.27f \geq \frac{d}{1-d} \tag{16}$$

Imagine that f, the chance of getting caught, is 50%. Then the condition simplifies to:

$$0.635 \geq \frac{d}{1-d}, \text{ or } d \leq \frac{0.635}{1.635} = 39\% \tag{17}$$

   In other words, even if criminals have only a 50% chance of being caught, if they retire with a 39% chance (or less), then their multiple "at bats" will catch up with them. In that case, the random criminal's lifetime chance of being caught and entered into the database will be 44%. And with a 44% chance that a random 1972 criminal will be in the 2004 database, the 125:1 odds that follow from a DNA database lead to a 99% chance that Puckett was indeed guilty.

---

the database contains the perpetrator and thus increase both the prior probability and the resulting posterior probability that the DNA match is the true source.

   104. An infinite series can be used because nearly all of the criminal class of 1972 will have either been caught or retired after thirty-two years. Thus, the probability of ending up in the database after thirty-two years is virtually identical to the probability of ending up in the database after infinity.

Significantly more than 50% of serious crimes are solved; f is closer to 0.61 than 0.50.[105] In addition, if a random criminal retires with a 39% chance, this says that the average criminal would commit 2.6 crimes before "retiring." This seems like a small number. If the average criminal commits 3.5 crimes before retiring, then $d = 0.29$. Plugging $f = 0.61$ and $d = 0.29$ into equation (14) leads to $p = 60\%$, well above the 44% threshold required to reach 99% confidence that Puckett was guilty.

This modeling approach to estimating the prior probability of database guilt—more precisely, the prior probability that someone in the database is the source of the crime scene DNA—has to our knowledge never before been used.[106] It is new in the sense that we start with the model of criminal behavior and use that to estimate the likelihood a criminal ends up in the database. Otherwise, when one looks at a database, such as the 338,711 names in the database used in *Puckett*,[107] there is no context. Simply dividing the database size by the population doesn't work for two reasons: it isn't clear what is the relevant population (San Francisco versus California), and it is clear that the database is far from a representative or random sample of the overall population.

In our behavior modeling approach, the prior probability doesn't seem to depend on the size of the database. It only depends on how often police catch criminals and how often criminals "retire." Of course, what matters for the prior is not the absolute number but the *fraction* of all criminals that end up in the database. That fraction and the database size are determined by the probability of being caught. If the police catch more criminals (f is higher), then we expect the database to be larger.[108] Moreover, as state and municipal databases have become less siloed, it might have become harder to retire out of them by moving to a different jurisdiction. Thus, if the database contains felons from all of California rather than just from San Francisco, then moving to Los Angeles is not enough to retire. The way in which the database is constructed will influence the chance that a convicted person will be entered into the database and that a criminal who is not caught ends up retiring. Over time, the FBI's national integration of local databases has increased D by effectively reducing d.

---

105. The national clearance rate for homicide has dropped significantly in the past fifty years from around 90% in 1960 to approximately 61% in 2006. Timothy G. Keel et al., *An Exploratory Analysis of Factors Affecting Homicide Investigations*, 13 HOMICIDE STUD. 50, 50 (2009).

106. It is, however, a standard economic approach to use observational results to estimate the underlying parameters of a model. *See, e.g.*, Ian Ayres, *Further Evidence of Discrimination in New Car Negotiations and Estimates of Its Cause*, 94 MICH. L. REV. 109, 128 (1995) (applying observational data on sellers' actual concessions to estimate which seller beliefs give rise to certain behavior); *see also* Orley Ashenfelter, *Estimating the Effect of Training Programs on Earnings*, 60 REV. ECON. & STAT. 47, 47 (1978) (using observed on-the-job training programs to parameterize an earnings model).

107. *See* Appellant's Opening Brief, *supra* note 70, at 44.

108. This assumes that the police catalog a DNA sample of each convicted person. This is a big "if" for cases in the 1970s, but it seems much more likely to be true today.

C. *Application to* Puckett

At the end of the day, do we think Puckett is guilty? It would seem so based on reasonable parameters of the model: (i) there is at least a 44% chance that the guilty party is in the database, and (ii) the DNA match provides a 125:1 updated likelihood. The end result is at least a 99% chance that Puckett is guilty.[109]

At this point, the prosecution might rest, but the case isn't over. The Bayesian approach has two critical and distinct roles to play in a trial. First, it provides the correct way to aggregate different information. Absent application of Bayesian updating, there is no coherent way for a juror to evaluate the relevance of statistical information such as random match probability. So far, we have used the Bayesian framework to present a coherent version of the prosecutor's case.

Once this framework has been adopted, there is a second value of the Bayesian approach: it should guide adversarial engagement on the important issues. For example, in *Puckett*, the defense would appropriately raise two significant objections.

The first is that Puckett does not seem to have followed the recidivism pattern assumed by our simple behavioral model. He was not a suspect in any other rape, assault, or murder other than the three he was convicted of in 1977. While Puckett was a serial rapist, those rapes all took place five years *after* Sylvester was murdered.[110] Thus, if guilty of the Sylvester murder, Puckett would have committed a rape and murder in 1972 followed by a five-year hiatus.[111] Furthermore, although he threatened all three victims, there were no stabbings or strangulations in any of the Puckett rape cases. Thus, Puckett would have gone from stabbing his victim to death to less violent attacks. Indeed, testimony from his three victims suggested that he was remorseful at the time and expressed concern for their well-being.[112]

The second objection is the presence of an alternative suspect whose existence was kept hidden from the jury. On November 8, 1972, "a mental patient

---

109. Although it led the *Puckett* jury to convict, the "conventional" approach of merely giving the jury the one-in-a-million random match probability does not provide a rational basis for factfinding.

110. Felch & Dolan, *supra* note 89.

111. It is possible that Puckett committed other rapes during that five-year interval and those crimes were never solved. However, there is no mention in the *Puckett* transcript (or elsewhere) of his DNA matching any other unsolved cases, and we can therefore infer that his DNA did not match any forensic evidence in any other cases.

112. *See* Appellant's Opening Brief, *supra* note 70, at 95-98. After raping Cynthia S., Puckett apologized, indicated that he wanted to be her friend, and offered her a job in the printing business. *Id*. at 97. Following the rape of Leslie B., Puckett drove Leslie back to Golden Gate Park (in her car), expressed remorse, and exited the vehicle. *Id*. at 98. Following the rape of Arlene M., Puckett "gave Arlene a hug, told her she was a neat person, and got out of the car." He knocked on the window and told her to "[m]ake sure she locked her door." *Id*. at 95-96 (alteration in original) (internal quotation marks omitted).

named Robert Baker escaped from Mt. Zion hospital in San Francisco."[113] His escape took place six weeks before the attack on Sylvester.[114] Shortly after Baker's escape, there was a sexual assault on Locksley Avenue, four blocks from Sylvester's apartment.[115] The rapist blindfolded and gagged the victim, declaring, "I can rape you now, or after you are dead."[116] The victim identified Baker as her assailant based on police photographs, but he was not arrested.[117]

Baker was also implicated in a series of stalking incidents that took place in September and October 1972. The police were called when a male stranger stalked a young mother and her nineteen-year-old nanny on the very same block where Diana Sylvester was murdered. On December 26, four days after the murder, the stalking victim, Lorraine Genigniani, identified the stalker as Baker from photographs and provided a license plate that matched a Volkswagen Bus registered to Baker.[118]

In early January 1973, Baker was apprehended, his Volkswagen was searched, and he appeared in a police lineup "probably related to" Diana Sylvester's rape and murder.[119] The results of that lineup are not known, and he was not arrested or charged.[120]

There is some question as to whether or not Baker fit the description of Sylvester's attacker. According to the defense brief,

> Sylvester's landlady, Helen Nigodoff, was the only person to give a description of the suspect. She provided two. In one statement, Nigodoff described the suspected perpetrator as "a WMA [white male adult], medium build and height, mustache, little beard, hair was dark brown." In the other, she described the person a little differently: "WMA, medium height, heavy set— chubby, curly br[ow]n hair, beard, mustache, clean cut appearance, d[a]rk clothing and jacket of denim-type." Like the suspect described by Nigodoff in both statements, Baker was an adult white male of medium height, with "dark hair" and "light beard growth."[121]

---

113. *Id.* at 74; *accord* Respondent's Brief, *supra* note 70, at 45.

114. Sylvester's murder occurred on December 22, 1972. Appellant's Opening Brief, *supra* note 70, at 1; Respondent's Brief, *supra* note 70, at 4; *see* Smith, *supra* note 70, at 74.

115. *See* Appellant's Opening Brief, *supra* note 70, at 3, 75; Respondent's Brief, *supra* note 70, at 45.

116. *See* Appellant's Opening Brief, *supra* note 70, at 76 (internal quotation marks omitted).

117. *See* Respondent's Brief, *supra* note 70, at 45. Baker was also identified by a neighbor who had seen Baker on the fire escape outside the victim's apartment shortly before the attack. Appellant's Opening Brief, *supra* note 70, at 75 n.32.

118. *See* Appellant's Opening Brief, *supra* note 70, at 74, 76-77, 78 & n.33; Respondent's Brief, *supra* note 70, at 45; Smith, *supra* note 70, at 74 ("According to police records, Baker also stalked a young girl and her nanny on Sylvester's block.").

119. *See* Appellant's Opening Brief, *supra* note 70, at 79; Respondent's Brief, *supra* note 70, at 46.

120. Respondent's Brief, *supra* note 70, at 46.

121. Appellant's Opening Brief, *supra* note 70, at 74-75 (alterations in original) (footnotes omitted) (citations omitted).

According to the State's brief, "the Locksley rapist was described as having a 'slender build,' weighing approximately 140 pounds. Helen Nigodoff, in contrast, described Diana Sylvester's killer as a 'heavy set—chubby.' And whereas Baker was 'very dirty and shabby,' Diana Sylvester's killer had a 'clean-cut appearance.'"[122] None of the twenty-six sets of fingerprints found in Sylvester's apartment matched Baker (nor did they match Puckett).[123] Among the evidence found in the investigation were a parking warning with the handwritten words "Robert the Artist" and type O blood, the same type as Diana Sylvester's[124] (and forty-four percent of the population[125]).

Baker died in 1978, and his DNA was not available in any database.[126] Thus, the real Baker is like our Mr. Baker on Eden: the one person whose DNA we'd like to test but cannot.[127] How should all this change our view of Puckett's guilt?

The alleged gap in Puckett's criminal activity is a challenge to our simple model of criminal behavior. That said, if he was Sylvester's killer, his retirement ended and his recidivism led to his being caught and included in the database. As for Baker, there is certainly some positive probability that he was Sylvester's killer. This could change our prior probability that the database concontains the guilty party. Based on the historical fraction of solved cases ($f = 61\%$) and 3.5 crimes committed per criminal ($d = 0.29$), we think a reasonable estimate for the prior probability is 60%.[128] The story of Baker helps us put a name on the other 40% (or a name on some part of that probability). Thus, it is reasonable to maintain the view that there is still a 60% chance that the killer is in the database and that Baker is the name that goes with some part of the other 40%.

A more conservative approach is to assign some explicit probability to Baker. If we think the probability Baker was the murderer is 30%,[129] then the most conservative and favorable assumption for Puckett is that the database and

---

122. Respondent's Brief, *supra* note 70, at 48 (citation omitted).

123. Felch & Dolan, *supra* note 89.

124. Respondent's Brief, *supra* note 70, at 46 (citations omitted).

125. *See Blood Types*, STAN. BLOOD CENTER, http://bloodcenter.stanford.edu/education /blood_types.html (last visited June 8, 2015).

126. *See* Michael Bobelian, *DNA's Dirty Little Secret*, WASH. MONTHLY (Mar./Apr. 2010), http://www.washingtonmonthly.com/features/2010/1003.bobelian.html.

127. It is possible that we could exclude or implicate Baker based on a DNA match of a sibling or child. We searched California and U.S. census records for siblings or offspring of Baker. Using records from the 1940 census, we were able to conclude that Baker had no older siblings, but we were unable to search for younger siblings or offspring as a result of privacy laws in California preventing the release of the information. *Personal Details for Arthur H Baker, "United States Census, 1940,"* FAMILYSEARCH, https://familysearch.org/pal: /MM9.1.1/K9DC-3BL (last visited June 8, 2015).

128. This estimate of p follows from calculations in equation (14).

129. Here we emphasize the point that Baker was a suspect but was not arrested. Thus, we infer that there must have been some alibi or evidence that led the police to change direction. For example, perhaps he was not identified in the lineup.

the other category continue to split up the remaining 70% in a 60:40 fashion. In other words, if it isn't Baker, then the remaining 70% still allows for people outside the database in the same proportion as it did before. That is, we ignore Baker as being part of the other category.[130] Following this approach, the database probability of containing the guilty party falls to 0.6 × 0.7 = 42%. In that case, the odds that Puckett is guilty are now (125 × 42):48, and so the posterior probability of guilt is 98.9%.[131] The inescapable logic of Bayes' rule is that even if there is another suspect with a 30% (prior) chance of being the guilty party, once we have a partial DNA match from Puckett, his chance of being the guilty party rises to a level (98.9%) which to our minds might still constitute proof beyond reasonable doubt.

VI. EMPIRICIZING THE ALIBI AND PRIOR PROBABILITIES

The application to *People v. Puckett* motivates a broader discussion of how to empirically assess the underlying parameters that influence the estimation of the posterior source probability. In particular, a central claim of this Article is that the appropriate application of Bayes' rule as laid out in equation (7) can lead courts and litigants to focus more explicitly on two variables that have to date been omitted from the adversarial process: a, the probability that non-sources in the database will have an alibi, and p, the prior probability that the source is in the database. This Part will suggest feasible methods for estimating these two probabilities and will suggest ways that expert testimony can be introduced with regard to the prior probability that the database includes the source without violating Federal Rule of Evidence 404's limitations regarding evidence of prior bad acts. We will first discuss how to parameterize a and p in large database tests (so-called "cold hit" or "trawl cases") and then discuss parameterization in very small database tests (so-called "confirmation cases") where the DNA of a single suspect or small number of suspects is tested against forensic evidence left at the crime scene.[132]

While equation (7) expresses the source probability as a function of the five variables, we focus here on a and p because there are more established conventions for ascertaining the values of the other three variables (M, r, and D). The science of DNA testing tells us the random match probability, r. D is simply the size of the database.[133] And the M in our model represents the number of unalibied matches, which is generally one for the cases of interest.

---

130. At the other extreme, we would assume that Baker comprises 30% of the 40% other category, and thus make no change to the 60% priors.

131. If the average criminal only commits two crimes, so that $d = 0.5$, then the prior falls to 38%, and the inclusion of Baker further reduces the odds Puckett is guilty to 45:1 or 97.8%.

132. *See generally* Thompson et al., *supra* note 62, at 6-7.

133. The alibi adjustment for the size of the database, however, has not to our knowledge ever been correctly implemented in a court proceeding. Thus, in *Puckett*, the govern-

A. *Large Database Trawls*

1. *Empiricizing the alibi probability*

The goal in assessing database alibis is to estimate the number of people in the database without alibis. In terms of equation (7), it is an attempt to estimate $D(1 - a)$. There is at least theoretically an analytically straightforward two-stage process for determining this quantity. The first, "categorical" stage is to exclude from the database those categories of individuals who belong to filterable groups that can algorithmically be eliminated from the database. The classic example would be eliminating women from a database search attempting to match on semen left at the crime scene. But, as described in our discussion of *Puckett*, categorical exclusion might also be applied on the basis of race or age.

After whittling down the effective size of the database through these categorical exclusions, the second, "idiosyncratic" stage would survey a random sample of the residual individuals to assess the proportion that have alibis of a sufficiently idiosyncratic type that they are not amenable to categorical exclusion. For example, as applied to *Puckett*, this second stage might entail a survey of 100 of the 8875 individuals (who remained after the age, sex, and race filters) to see what proportion would have a credible alibi for being at a different place at the time of the crime. If twenty-five percent had credible alibis, then the net database size should be further reduced by this probabilistic amount. The quantum $E[M]$ for the expected number of unalibied matches is found by multiplying the unadjusted database size by both the random match probability and the probability of not having an alibi: $E[M] = rD(1 - a)$. The end result is the same whether one first looks at the number of matches (rD) and then reduces that number to those without alibis ($1 - a$) or one starts with a reduced database of those without alibis ($D(1 - a)$) and then looks for matches (r). And one can break up the no-alibi probability into two parts, $(1 - a) = (1 - a_1) \times (1 - a_2)$, where $(1 - a_1)$ is the fraction of the database that have the correct demographics and $(1 - a_2)$ is the fraction of those with the correct demographics who have no idiosyncratic alibi.

While it is relatively simple to calculate the fraction of a database with categorical or demographic exclusions, conducting an investigation to determine the idiosyncratic-alibi probability might be infeasible given the limited resources of litigants. But no one promised that applying Bayes' rule would be costless. One might imagine developing general evidence-based ranges of a for certain types of crimes. For example, we might learn that the alibi probability naturally diminishes over time as individuals lose track of their whereabouts at particular moments in the distant past. These general probabilities might or might not be sufficiently credible to use as an input in estimating the posterior source probability. Our point here is not to authoritatively resolve all these dis-

---

ment expert testified on the basis of a 338,711 database size, when net of alibis the adjusted database ($D(1 - a)$) was likely less than 10,000.

putes, but instead to show how our analysis can help courts and litigants better engage what should be the central evidentiary issues.

2. *Empiricizing the prior probability*

Courts have been particularly reluctant to introduce evidence about the prior probability of an individual defendant's being the source of forensic DNA[134] because the very process of constructing a prior probability seems inconsistent with the notion of a fair trial. Here we shift the question to how to ascertain the prior probability that the database contains the source of the crime scene evidence. After describing alternative methods for estimating this probability, we take on the difficult question of whether such evidence should be introduced in a criminal trial.

One method of empiricizing this prior probability is to make inferences based upon a structural model of how individuals become part of the database. Structural models can produce markedly different estimates of the prior probability. Consider the following examples:

1. If the state executes or otherwise incapacitates all convicted felons, and the database is solely comprised of convicted felons, then if the crime is committed after the felon enters the database, the prior probability would be zero.
2. If a murder occurs on a prison cell block, where the only possible perpetrators are the 200 felons currently incarcerated, and the database is solely comprised of convicted felons, then the prior probability could be one hundred percent or very close to it.
3. If a jurisdiction randomly selects a proportion of the population to be included in the DNA database, then the prior probability would just be that proportion.[135]

In discussing *Puckett*, we developed a simple structural model that turned on the clearance rate and the likelihood that a criminal would retire from committing crimes in that jurisdiction.[136] The structural model has the advantage of making clear how assumptions about the world lead to estimates of the prior. But, as we show below, one can also estimate the database prior directly.

An alternative to structural modeling of the prior is to collect statistical evidence on what we will call the "adjusted match probability." This approach starts by examining database trawls of similar crimes to assess the frequency with which these similar trawls produced an unalibied match. Imagine, for example, that there have been 100 similar rape cases where forensic evidence has

---

134. *See* Thompson, *supra* note 12, at 10-12.

135. Even here, there is a question of what is the appropriate population. Should it be the population of San Francisco, California, the United States, or all of Earth?

136. The clearance rate is a published statistic. We can make some estimates for retirement rates based on the fraction of defendants who have a prior conviction.

been tested. Suppose 30% of these trawls produced a match without an airtight alibi.[137] This 30% number is the starting point for a direct estimate of the prior.

To estimate the prior probability for a similar crime, one would adjust downward this raw match proportion to take into account both the possibility that (i) some of these matches might by chance have been matches with non-source individuals and (ii) the sample of 100 trawls might by chance have been overrepresentative of the larger population. In the next Part's discussion of small database trawls, we will provide specifics on how this might be accomplished.[138] With a 30% hit rate based on 100 trawls, we show that with 99% confidence one can reject the conclusion that the true source probability is anything below 19.7%.[139] This is a conservative adjusted match rate that should be used as the prior database source probability. This approach can be seen as using the average posterior source probability found in similar cases, with adjustments made due to the small sample size and the potential for random matches.

Of course, there will be and should be reasonable disagreement about what constitutes a similar crime.[140] Is rape similar to combined murder and rape? Similarity would have to be with regard to a host of factors—including not just the crime type but also the modus operandi[141] and the characteristics of the defendant.[142] But, as an empirically contingent matter, it might be possible to develop general ranges of probabilities that a match will occur.[143] A sensible way forward would be to derive alternative priors based on alternative assumptions of what constitutes similar crimes as well as on plausible structural models, and then see if the defendant's source probability is sufficiently high even after combining the likelihood ratio with the most conservative (i.e., lowest) probability estimate within this range.

B.   *Small Database Trawls*

The analysis above is done under the stylized assumption that everyone in the database has the same prior probability of having been at the crime scene.

---

137. An example of an airtight alibi is that the perpetrator was male but the matching individual was female.

138. *See infra* notes 147-48.

139. *See infra* notes 147-48. The 19.7% figure is derived by applying the formulas in those two notes.

140. *See, e.g.*, Jason L. McCandless, Note, *Prior Bad Acts and Two Bad Rules: The Fundamental Unfairness of Federal Rules of Evidence 413 and 414*, 5 WM. & MARY BILL RTS. J. 689, 707 (1997).

141. It might, for example, be possible that the prevalence of stranger rapists in a dataset deviates from the prevalence of acquaintance rapists.

142. It might, for example, be possible that African Americans, because of disparities in policing, are overrepresented in the database.

143. With a large enough set of test trawls, it might be possible to back out an implicit prior probability by calculating how much more often one observes matches in the aggregate than one would expect by chance.

Of course, some individuals will have a much higher prior probability than others. Take the case where a woman, who had a documented history of being a victim of spousal battery, is found murdered. There is no evidence or witness testimony linking the husband to the crime, but he also doesn't have an airtight alibi. Based on other cases where a battered wife is found murdered with incriminating DNA evidence left at the crime scene, what is the prior probability that the husband is the source of the forensic evidence?[144]

In this case, the database is of size $D = 1$. Testing the DNA of a single suspect or a small number of suspects is an example of a confirmation test. In such settings, it is straightforward to estimate the underlying parameters, a and p—especially the alibi probability. Because of the small size of the dataset, we can check to see if the husband has an alibi. If so, he is excluded. If not, then $a = 0$ and we can proceed to the DNA testing. Thus, in the foregoing husband example: $D = 1$, $a = 0$, $E[M] = r$, which we will assume to be one in a million, and $M = 1$ (because the question would be moot if the husband didn't match).[145]

The only remaining question is, what is the prior? That is, what is the probability that a database consisting of husbands without alibis will include the source?

We again suggest that the prior can be inferred from adjusting the match rate in similar confirmation cases. Sticking with the husband example, imagine that we find a hundred other examples where the DNA of husbands without alibis is tested to see if they were the source of incriminating DNA left at the crime scene where their battered spouse was murdered. And imagine we ob-

---

144. In the highly publicized O.J. Simpson murder trial, the defense attempted to exclude evidence that Simpson battered his wife before she was murdered. *See* ALAN M. DERSHOWITZ, REASONABLE DOUBTS: THE O.J. SIMPSON CASE AND THE CRIMINAL JUSTICE SYSTEM 104 (1996). Alan Dershowitz, working for the defense, argued that evidence of battery was irrelevant because less than one-tenth of one percent of batterers murder their spouses. *Id.* Dershowitz was criticized for using the wrong prior probability. Gerd Gigerenzer, *Ecological Intelligence: An Adaptation for Frequencies*, *in* THE EVOLUTION OF MIND 9, 18-20 (Denise Dellarosa Cummins & Colin Allen eds., 1998). The correct prior probability is the probability that the husband is the murderer given that the husband was a batterer *and the woman was murdered*. *See* I.J. Good, Letter to the Editor, *When Batterer Turns Murderer*, 375 NATURE 541 (1995); Jon F. Merz & Jonathan P. Caulkins, *Propensity to Abuse—Propensity to Murder?*, 8 CHANCE, no. 2, 1995, at 14; *see also* I.J. Good, Letter to the Editor, *When Batterer Becomes Murderer*, 381 NATURE 481 (1996). Dershowitz's mistake is recounted in numerous statistics textbooks and commentaries. *See, e.g.*, HENK TIJMS, UNDERSTANDING PROBABILITY 273-75 (3d ed. 2012); Gigerenzer, *supra*, at 18-20. For a discussion of the evidentiary issues surrounding the admission of the battery evidence, see Roger C. Park, *Character Evidence Issues in the O.J. Simpson Case—or, Rationales of the Character Evidence Ban, with Illustrations from the Simpson Case*, 67 U. COLO. L. REV. 747, 748-54 (1996). *Cf.* Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859, 860-61 (1996) (applying Bayesian statistics to the probability evidence in the Simpson case).

145. If the DNA sample is more than a fragment, then r would be less than one in a quadrillion and the prior probability is hardly relevant.

serve that in these hundred tests, the husband's DNA matched the forensic DNA thirty times.[146]

It might be tempting to infer that thirty percent of the time the husband was the source of the forensic DNA. But this inference would ignore that some of the matches might have occurred by chance with husbands who were not the source. While it is impossible to infer with certainty in any individual case whether a matching husband was the true source or a nonsource who matched by chance, it is possible in aggregate to infer the probability that the true source was included in these hundred databases of size one.[147] The probability that the source was included in these databases in the hundred similar cases can be used as the prior probability that a database consisting of a husband without an alibi will include the source of the forensic DNA. For the parameter values (r is one in a million, μ is 0.30), we can reject with 99% confidence that the true probability of a match is anything below 19.7%.[148] This lower-bound estimate would provide a conservative method of estimating the prior that relies on the kinds of evidence that could be preserved by testing labs.

While the foregoing example assumes that the marital relationship is the sole salient factor in determining similar cases, we emphasize that litigants and their experts should consider alternative indicators of similarity as well. We could imagine priors that were derived from multivariate probit or logit regressions that began by predicting the unadjusted matching rate from a host of crime and suspect characteristics. This probability could then be adjusted downward to take into account the confidence interval and the likelihood of nonsource matching.

---

146. In these cases, the husband may have a legitimate reason to have his DNA present and so being the source is not the same as being guilty. Also, the husband might be guilty even when there is no incriminating DNA left at the scene.

147. If we let the proportion of tests that uncovered matches be μ, then we can infer p by realizing that μ equals the probability that the source is in the dataset (p) plus the probability that the source is not in the dataset times the probability that a nonsource will match, $((1 - p)r)$:

$$\mu = p + (1 - p)r \Rightarrow p = (\mu - r)/(1 - r) \qquad (\text{F1})$$

For the parameter values assumed above (r is one in a million, μ is 0.30), p is 29.99%.

148. The Wilson score interval provides a tool for calculating a confidence interval when observing draws from a binomial distribution. To find a lower bound on the 99% confidence interval, one looks for the solution to:

$$\mu_l + 2.58\sqrt{\mu_l(1 - \mu_l)/n} = \mu$$

where μ is the observed frequency, n is the sample size, and 2.58 is the number of standard deviations associated with 99% confidence. With $n = 100$, $\mu = 0.30$, application of the quadratic formula leads to $\mu_l = 0.197$ or 19.7%. Using this lower bound in equation (F1) above leads to a conservative estimate of $p = 19.6999\%$. The correction for random match probability has almost no effect, but the correction for small sample size is significant. For an explanation of the Wilson score, see Sean Wallis, *Binomial Confidence Intervals and Contingency Tests: Mathematical Fundamentals and the Evaluation of Alternative Methods*, 20 J. QUANTITATIVE LINGUISTICS 178, 183-87 (2013).

Stepping back, this Part has shown practical means of estimating the alibi and the prior probability variables that should play an essential role in estimating the posterior probability that the defendant was the source of the forensic DNA. Our suggested approaches do not turn advocates or jurors into potted plants who must blindly accept as gospel expert estimates. Rather, by showing the range of choices and approaches on which experts might base their estimates, we have provided a guide for enlightened advocacy and inference.

VII. ADMISSIBILITY

Our goal of this Part is to suggest specific ways an expert might present his or her opinion under existing law and when existing evidentiary rules should change to accommodate a more coherent factfinding process. We assess the admissibility of posterior source probability and its antecedents—especially the admissibility of the prior, the database source probability that we have argued will often best be captured by an adjusted match rate of similar crimes.

The general problem is that courts are reluctant to admit evidence of prior bad acts as evidence of a defendant's guilt. Federal Rule of Evidence 404 provides that "[e]vidence of a crime, wrong, or other act is not admissible to prove a person's character in order to show that on a particular occasion the person acted in accordance with the character," but "may be admissible for another purpose, such as proving motive, opportunity, intent, preparation, plan, knowledge, identity, absence of mistake, or lack of accident."[149] For example, parties may not introduce "evidence of a violent disposition to prove that the person was the aggressor in an affray, or evidence of honesty in disproof of a charge of theft."[150] The rule thus "implements the philosophy that a defendant should not be convicted because he is an unsavory person, nor because of past misdeeds, but only because of his guilt of the particular crime charged."[151]

Although the adjusted match probability is derived from database trawls of similar criminal cases, it is not inadmissible character evidence. Unlike character evidence, which is used to establish that the defendant acted in accordance with her propensity to act criminally, the adjusted match probability is the likelihood that the *database* is guilty. This purpose falls into at least one of the permitted uses—identity.[152] No reasonable juror would infer from crimes

---

149. FED. R. EVID. 404(b). In *Puckett*, the defense unsuccessfully attempted to exclude evidence of the defendant's sex crimes under section 1101 of the California Evidence Code, the state equivalent of Rule 404. *See supra* note 84.

150. FED. R. EVID. 404 advisory committee's note.

151. 1 CHRISTOPHER B. MUELLER & LAIRD C. KIRKPATRICK, FEDERAL EVIDENCE § 4:21, at 677 (4th ed. 2013).

152. *See* FED. R. EVID. 404(b)(2) ("[E]vidence [of a crime, wrong, or other act] may be admissible for another purpose, such as proving motive, opportunity, intent, preparation, plan, knowledge, *identity*, absence of mistake, or lack of accident." (emphasis added)); *see also* CAL. EVID. CODE § 1101(b) (West 2015) ("Nothing in this section prohibits the admission of evidence that a person committed a crime, civil wrong, or other act when relevant to

committed by others that the defendant acted in accordance with her propensity to act criminally. The emphasis is not on the prior probability that the defendant who matched is guilty, or even the prior probability that the defendant is the source, but rather on the likelihood that the database includes the source regardless of who the source might be. This database source probability thus provides one measure of reliability: how reliably a database trawl uncovers the source of the crime scene evidence.

Rule 404 is not, however, the only evidentiary rule that must be satisfied. To be admitted, the proposed probability evidence must also be consistent with Rules 104(a), 702, 703, and 403, or their state equivalents.[153] Rules 104(a) and 702 govern the admission of expert testimony. Rule 104(a) authorizes a trial court to decide "whether a witness is qualified . . . or evidence is admissible."[154] Rule 702 allows the court to admit expert testimony if

> (a) the expert's . . . knowledge will help the trier of fact to understand the evidence or to determine a fact in issue;
>
> (b) the testimony is based on sufficient facts or data;
>
> (c) the testimony is the product of reliable principles and methods; and
>
> (d) the expert has reliably applied the principles and methods to the facts of the case.[155]

Courts routinely have admitted match probabilities and other DNA statistical evidence under this constellation of rules.[156]

Rule 703 is particularly relevant because it expressly allows an expert to give an opinion based upon otherwise inadmissible underlying facts or data if "experts in the particular field would reasonably rely on those kinds of facts or data in forming an opinion on the subject."[157] The empirical estimates of the parameters underlying Bayes' rule are just those kinds of facts upon which statistical experts would rely in forming an opinion as to the posterior source

---

prove some fact (such as . . . *identity* . . .) other than his or her disposition to commit such an act." (emphasis added)).

153. For a discussion of the legal framework governing the admission of scientific evidence, see LAWRENCE KOBILINSKY ET AL., DNA: FORENSIC AND LEGAL APPLICATIONS 200-07 (2005). *See also* FED. R. EVID. 104(a) (concerning whether "a witness is qualified, a privilege exists, or evidence is admissible"); *id.* 402 (concerning the general admissibility of relevant evidence); *id.* 403 (concerning the admissibility of prejudicial but probative evidence); *id.* 702 (concerning the admissibility of opinion testimony); *id.* 703 (concerning the admissibility of expert witness testimony); Daubert v. Merrell Dow Pharm., Inc., 509 U.S. 579, 582, 597 (1993) (establishing the standard by which a trial court should assess whether an expert's scientific testimony is admissible).

154. FED. R. EVID. 104(a).

155. *Id.* 702.

156. *See* KOBILINSKY ET AL., *supra* note 153, at 207. *Puckett* illustrates the similarities between the Federal Rules of Evidence and their state equivalents as applied to DNA statistical evidence. The prosecution at trial blocked the defense's attempt to admit such evidence under section 352 of the California Evidence Code, the state equivalent of Rule 403, arguing that the admission of the defendant's DNA profile would mislead the jury. *See* Respondent's Brief, *supra* note 70, at 7-10, 17-19.

157. FED. R. EVID. 703.

probability. Therefore, Rule 703 might create an opportunity for experts to testify as to the bottom-line (posterior) source probability without referring to the underlying prior database source probability or how it was estimated. Thus, by exploiting Rule 703, courts might be able to avoid concerns under Rule 403 about whether the probative value of admitting estimates of the prior database source probability "is substantially outweighed by a danger of . . . unfair prejudice, confusing the issues, misleading the jury, undue delay, wasting time, or needlessly presenting cumulative evidence."[158]

How likely are courts to admit adjusted match probabilities derived from database trawls of similar crimes under these rules? Analysis of judicial opinions suggests that although courts do not agree over whether prior probability evidence is admissible, they are more likely to admit it in the presence of certain factors. Jonathan Koehler argues that courts are more willing to admit prior probability evidence when the evidence is offered to rebut an argument that a given outcome happened by chance,[159] and when the reference class from which the probability is derived is narrow and specific to the focal case.[160]

In the case of adjusted match probabilities, at least one of these factors is present. Adjusted match probabilities, like other match probabilities, provide jurors with the information needed to determine the likelihood that the defendant is the source of the DNA and the match did not happen by chance.[161] Accordingly, whether such a probability is admissible may depend on how similar the instant case is to the class of reference crimes. The more similar the current crime is to the reference class of crimes, the more likely the court will be to admit the proposed evidence.

It is difficult to predict how courts will resolve the question of similarity.[162] Outside the sentencing context, courts rarely address statistical evidence related to other crimes.[163] However, changes in courts' approaches to similarity with respect to DNA evidence provide some reason for predicting that adjusted match probabilities are increasingly likely to be admissible. When random

158. *Id*. 403. The use of expert testimony to obviate the need to admit estimates of the prior probability raises separate questions of fairness. Many defendants will not have the resources to hire opposing experts.

159. *See* Jonathan J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?*, 42 JURIMETRICS 373, 388-90 (2002). Koehler derives his conclusions from a comprehensive study of more than 100 federal and state opinions discussing the admission of base rate evidence. *Id*. at 379. We could apply our base rate approach to estimate a prior probability that the court will admit evidence about prior probabilities.

160. *See id*. at 390-91.

161. *See id*. at 388-89. Similarly, courts have accepted probability evidence to rebut chance arguments in cases involving suspicious deaths and test scores. *See id*. at 389-90.

162. *See id*. at 401 ("Admittedly, the identification of patterns in court rulings related to the perceived relevance of base rates is part science and part art. Courts that consider the relevance of statistical evidence have different philosophies, are influenced by a host of situation-specific variables, and occasionally make rulings that might go the other way.").

163. *See, e.g.*, Dorsey v. State, 350 A.2d 665, 668 (Md. 1976) (holding that a detective's testimony that seventy-five to eighty percent of his arrests led to convictions was inadmissible).

match probabilities, described in Part II, were introduced into the courtroom, defendants challenged their admission on the grounds that the populations from which they were derived did not account for genetic variations across ethnic groups.[164] However, these challenges proved unsuccessful, and random match probabilities are widely accepted by courts.[165] Indeed, some courts have held that it is incoherent—and error—to admit DNA evidence without accompanying statistical evidence.[166]

However courts decide these issues, we believe that as a normative matter, admitting the adjusted match probability would not be unfairly prejudicial. First, Bayesian logic prevents the adjusted match ratio from dominating jurors' calculations of the posterior odds. The adjusted match probability only establishes a prior probability. The prosecution must still establish the likelihood ratio based on the matching evidence in the specific case. Moreover, even if the prosecution shows that there is a high prior that the database includes the source, the prosecution must further show, based on the matching evidence, a high likelihood that the defendant is the source, and that the defendant is guilty. Again, this demonstration will depend on the facts of the case. Second, defendants would have the opportunity to challenge the admission of adjusted match probabilities derived from insufficiently similar reference classes. Under the scheme we contemplate, similarity would be a contested issue. A defendant might move to exclude an adjusted match probability derived from database matches in cases involving crimes other than the charged crime. Or she might argue that such databases were insufficiently similar because of geographic or demographic differences. At trial, parties could introduce expert testimony regarding the effects of these differences on the prior probability.

A judicial demand for individualized evidence should not at the end of the day preclude the admissibility of database source probabilities. Because the Bayesian posterior of equation (7) is literally the product of an individualized likelihood ratio and a nonindividualized prior, the bottom-line posterior source probability is individualized.

Finally, even if courts ruled that adjusted match probabilities were not admissible, the effect of these rulings could be mitigated through testimony regarding the effect that varying the prior probability would have on the posterior probability. Experts could present a chart mapping different prior probabilities to posterior probabilities.[167] Showing how the posterior depends on particular

---

164. *See* Koehler, *supra* note 159, at 393.

165. *See id.* at 388 ("Almost all courts find [random match] probabilities to be relevant and admissible."); *see also supra* note 12 and accompanying text.

166. *See, e.g.,* Hull v. State, 687 So. 2d 708, 728 (Miss. 1996).

167. *See* KAYE ET AL., *supra* note 33, § 14.3.2(c), at 652. This approach is consistent with appellate court holdings. David Kaye and his coauthors note that

> [a]lthough American courts are divided as to the admissibility of posterior probabilities derived through Bayes' rule with undisclosed prior probabilities—particularly in criminal cases—no appellate court in this country has held that it is error to permit testimony about the impact of the trace evidence on the prior probability.

priors would allow experts to explain the minimum prior belief needed to produce a posterior belief of at least 99%. Under this approach, jurors would not be required to accept any single figure for the prior probability and would be left to determine the prior from other parts of the record. For matches on multiple loci, the likelihood ratio will be so high—on the order of one in a quadrillion—that a prior probability would have to be less than one in a trillion to produce a posterior with less than a 99.9% posterior chance that the defendant was the source of the crime scene evidence.

While limiting an expert to just reporting on the likelihood ratio and instructing the jury on what they might demand as a minimally necessary prior probability of database guilt avoids many of the current obstacles to admissibility, we think that courts and litigants should embrace full-throated Bayesian analysis.

To those who object to our proposal, we challenge you to provide a coherent alternative. The status quo fails that test. Telling a jury a likelihood ratio without a prior is akin to telling someone how many eggs to include in a cake recipe without telling them how much flour to use or the serving size. The 125:1 likelihood ratio that comes out of the Puckett partial DNA match is damning evidence if a juror thinks that there is a forty-four percent chance the guilty party is in the database. But it is evidence of a very weak case if the prior chance is below one in two hundred.

Moreover, we think it is irresponsible to believe that jurors can somehow figure out all on their own how to estimate priors and how to aggregate prior beliefs with information from database matches. This is asking jurors to rediscover Bayes' rule without any help. Worse still, if jurors are denied knowledge of how the match was found, then they don't have any basis on which to form a prior. Somehow courts think it is better for jurors to rely on uniformed and ignorant prejudices than to risk confusing them with the relevant data to make an informed decision.

What is surprising to us is that the statistical experts seem to agree with the view that the jurors should be left to calculate and update priors on their own. Balding and Donnelly write: "The assessment of the other evidence is a matter for the jury, not an expert witness. Consequently it is inappropriate for an expert to express a view on whether the defendant was the source of the crime sample . . . ."[168] A.P. Dawid writes: "I agree . . . that it should be left to the juror to assess and incorporate his or her own prior probabilities rather than have these supplied, explicitly or implicitly, by expert witnesses . . . ."[169]

Id.

168. *See* Balding & Donnelly, *How Convincing Is DNA Evidence?*, *supra* note 58, at 286.

169. A.P. Dawid, Letter to the Editor, *Comment on Stockmarr's "Likelihood Ratios for Evaluating DNA Evidence when the Suspect Is Found Through a Database Search,"* 57 BIOMETRICS 976, 978 (2001).

In other words, the experts provide the jury with the 125:1 likelihood figure, and then the jurors are left on their own to figure out what this means. Without incorporating a prior probability, it means nothing. As Louis Kaplow writes, "It is as if one was asked to choose the rectangle with the greater area, but in so doing to ignore the length of the rectangles under consideration."[170]

Put yourself in the jury box. You are told that Puckett is a thrice-convicted rapist whose DNA matches that found at the crime scene and the chance of a random match in the relevant database is 1/125, or 0.8%.[171] This might sound much more convincing than it really is.

If the jurors accept the 125:1 figure, the chance that the observed match is the true source is $125p/(125p + 1 - p)$, where p is the prior probability of a database match. But where do the priors come from? Are jurors supposed to build and fit a model for the prior chance that a random person in the California felon database with the right demographics and no alibi is guilty of a specific murder? How are they to understand the size of the relevant database without expert information about the composition of that database?

As we've seen, calculating a prior can be just as complicated as calculating the 125:1 ratio used to update the prior. And since the two ratios are equally essential in the calculation, we think experts should help guide the jurors in estimating both. As suggested earlier, it would be useful to provide jurors with a lower bound on prior beliefs needed to convict. Instead of asking jurors where they end up, we ask jurors where they start, or at least where they start after hearing the non-DNA evidence but before they know about a match. The expert can help the jury appreciate the minimum threshold that they need to be above in order to reach a conclusion.

As we argue that priors matter, perhaps they might matter too much. Another critique of our approach is that it suggests that someone might be convicted based on a high prior without any additional evidence. This is the classic Blue Bus problem.[172] In fact, Hertz was unable to obtain summary judgment to escape liability for an accident in which the guilty party was identified as driving a yellow Hertz truck.[173] Hertz didn't own all yellow trucks with its name on them, but did own ninety percent of such trucks, and that was enough to prevent summary judgment.[174]

Allow us to consider two cases. In case 1, all we know is that the guilty party is a blue bus and company *A* owns 80% of the blue buses. This is the

170. Louis Kaplow, *Burden of Proof*, 121 YALE L.J. 738, 798 (2012).

171. The prosecution might try to claim that the chance of a random match is one in a million, while the defense might try to counter that the chance of a random match is one in three. But experts would point out that the relevant database is neither one person nor the entire list of 338,711 felons. The relevant database is closer to 8875, so the chance of at least one random match is $1 - (1{,}099{,}999/1{,}100{,}000)^{8875} = 0.8\%$.

172. The Blue Bus problem is derived from *Smith v. Rapid Transit, Inc.*, 58 N.E.2d 754 (Mass. 1945). *See* Tribe, *supra* note 13, at 1340, 1341 & n.37.

173. *See* Kaminsky v. Hertz Corp., 288 N.W.2d 426 (Mich. Ct. App. 1979).

174. *See id.* at 427.

classic Blue Bus problem. In case 2, the victim is able to identify the last digit of the license plate as a 3. In this scenario, company *A* and company *B* each have one bus. The license plate from company *A*'s bus ends with a 3, and the plate of company *B*'s bus ends with an 8. If the eyewitness is correct, then company *A* is surely the guilty party. But testing shows that, in similar visual conditions, the witness is only able to correctly identify a 3 as a 3 80% of the time; the other 20% of the time he sees a 3 as an 8. Similarly, he sees an 8 as an 8 80% of the time and the other 20% of the time misidentifies it as a 3.

Thus "seeing" a 3 would lead to 4:1 odds that the truth is a 3. In this second case the prior is 1:1 so that the posterior is 4:1. And this endpoint is the exact same as the starting point for company *A* in the original Blue Bus problem. In both cases the chance that company *A* is the guilty party is 80%.

Is one 80% better than the other? Is there any more reason to find liability because of strong (but imperfect) eyewitness evidence combined with a weak prior than to base the decision on equally strong (but imperfect) initial beliefs combined with no additional data?

We think not. Our view is that data are data. Prior data are just as important as data that allow us to update our beliefs. What ultimately matters is where we end up,[175] and that we arrived at that destination via a path that employs sound logic and reasoning.

CONCLUSION

In the 2012 presidential election, Nate Silver caused a stir by correctly predicting the winner of all fifty states and the District of Columbia in the general election.[176] Beyond accuracy, Silver's larger impact has been in changing the central polling metric and improving the way that metric is calculated. Before Silver, the discussion of preelection polling focused on the extent to which one candidate led another candidate in the polls.[177] Silver's approach instead fo-

175. As Steve Salop suggested to us, the two cases will lead to different incentives for accident mitigation. In case 1, company *B* has no incentive to reduce the number of accidents, while company *A* bears the full costs of all its accidents and thus has the full incentive to engage in accident mitigation. In case 2, each company only has an 80% chance of being found liable for its accidents (as the eyewitness may read the license plate incorrectly), and thus each company only has 80% of the full incentive. Note that in both cases, the incentives are 80% but are allocated differently. In case 1, 80% of the market has a 100% incentive, while in case 2, 100% of the market has an 80% incentive.

176. Daniel Terdiman, *Obama's Win a Big Vindication for Nate Silver, King of the Quants*, CNET (Nov. 6, 2012, 11:50 PM PST), http://www.cnet.com/news/obamas-win-a-big-vindication-for-nate-silver-king-of-the-quants.

177. *See, e.g.*, Rich Jaroslovsky, *Clinton Holds 10-Point Lead in Poll but Bush's Approval Rate Improves*, WALL ST. J., Sept. 17, 1992, at A18; Richard L. Berke & Janet Elder, *Poll Shows Either Candidate Within Reach of Victory*, N.Y. TIMES (Nov. 6, 2000), http://www.nytimes.com/2000/11/06/us/the-2000-campaign-the-polls-poll-shows-either-candidate-within-reach-of-victory.html; Paul Harris, *Race for President Back in Balance*, GUARDIAN (Oct. 9, 2014, 8:13 PM EDT), http://www.theguardian.com/world/2004/oct/10/uselections2004.usa1; Jim Rutenberg, *After First Debate, Both Sides Work Hard to Sway Perceptions*,

cused on what we care about: the probability a candidate would win the election.[178] Before Silver, pundits would say: Obama leads Romney 52% to 48% with a margin of error of plus or minus 2%. After Silver, the same evidence can be described as a 97.5% chance that Obama will win the election.[179]

Silver's method of estimating these probabilities of victory is explicitly motivated by the Bayesian logic of continually updating predictions based on the quality of additional evidence.[180] Like a good Bayesian, Silver gives different weights to different types of information based on testable evidence of the information's likely accuracy.

Our approach in this Article mirrors Silver's contributions. We advocate changing the central statistic of interest. Just as the public cares most about a candidate's chances of winning, the jury should primarily care about the probability that the defendant was the source of DNA left at the crime scene. And, like Silver, we advocate that the method of estimating this probability be explicitly Bayesian.

A move toward Bayesianism will not mean the end of adversarialism. Rather, we have given advocates clear grounds on which to assess and challenge the quality of the evidence. While Bayesian logic tells us to deflate the nominal size of the database to exclude those with alibis, there are consequential choices in determining what methods are used to implement categorical or probabilistic exclusion.[181] And while Bayes' rule tells us that the prior probability of "database guilt" plays a central role in calculating the defendant's posterior source probability, litigants have not sufficiently engaged on the radically different priors that can be produced from estimates based on different structural or survey methods. Defense attorneys may be concerned that an explicit embrace will slam the door on all efforts to defend defendants with DNA matches. But credible evidence of a low prior probability that the database includes the DNA source combined with a partial match may lead to more reasonable doubt. It can also lead to more reasonable certainty than the current incoherent practice—seen in cases such as *Puckett*—of telling jurors, "The defendant's DNA matched that found at the scene, we won't tell you how the match was found, but the chance of a random match is less than one in a million."

N.Y. TIMES (Oct. 4, 2004), http://www.nytimes.com/2004/10/04/politics/campaign/04spin.html.

178. *See* Terdiman, *supra* note 176.

179. In fact, Silver succeeded in putting into practice a suggestion one of the authors made in 2007. "Instead of reporting the margin of error, reporters should start telling people something that they intuitively understand, the 'probability of leading.'" IAN AYRES, SUPER CRUNCHERS 202-04 (2007).

180. NATE SILVER, THE SIGNAL AND THE NOISE: WHY SO MANY PREDICTIONS FAIL— BUT SOME DON'T 248 (2012); *see also* Gary Marcus & Ernest Davis, *What Nate Silver Gets Wrong*, NEW YORKER (Jan. 25, 2013), http://www.newyorker.com/books/page-turner/what-nate-silver-gets-wrong (noting that Silver is a champion of Bayesian statistics, but arguing that he overstates the usefulness of Bayesian statistics and that his discussion of alternative, non-Bayesian methods is "dismissive, incomplete, and misleading").

181. *See supra* Part V.A (discussing alternative ways of implementing alibi exclusions).