

2-12-2012

The Evidence of Things Not Seen: Non-Matches as Evidence of Innocence

James S. Liebman

Columbia University School of Law, jl Liebman@law.columbia.edu

Follow this and additional works at: http://lsr.nellco.org/columbia_pllt

 Part of the [Criminal Law Commons](#), [Criminal Procedure Commons](#), [Evidence Commons](#), [Law and Society Commons](#), [Psychology and Psychiatry Commons](#), [Public Law and Legal Theory Commons](#), and the [Science and Technology Commons](#)

Recommended Citation

Liebman, James S., "The Evidence of Things Not Seen: Non-Matches as Evidence of Innocence" (2012). *Columbia Public Law & Legal Theory Working Papers*. Paper 9201.

http://lsr.nellco.org/columbia_pllt/9201

This Article is brought to you for free and open access by the Columbia Law School at NELCO Legal Scholarship Repository. It has been accepted for inclusion in Columbia Public Law & Legal Theory Working Papers by an authorized administrator of NELCO Legal Scholarship Repository. For more information, please contact tracy.thompson@nellco.org.

© 2012 James S. Liebman. Do not quote, cite or distribute without permission of the lead author.

THE EVIDENCE OF THINGS NOT SEEN:¹
NON-MATCHES AS EVIDENCE OF INNOCENCE

James S. Liebman,² Shawn Blackburn,³ David Mattern,⁴ Jonathan Wainor⁵

Exonerations famously reveal that eyewitness identifications, confessions and other “direct” evidence can be false, though police and jurors greatly value them. Exonerations also reveal that “circumstantial” non-matches between culprit and defendant—e.g., a non-matching aspect of an eyewitness’s description or a loose button at a crime scene—can be telling. Although non-match clues seem uninteresting because they are easily explained away, they often turn out to match the real culprit when he or she is eventually caught. This article uses “non-exclusionary non-matches” and their seeming polar opposite, inculpatory DNA, to show that: (1) all evidence of identity derives its power from the aggregation of individually uninteresting matches or non-matches, but (2) our minds and criminal procedures conspire to hide this fact when they contemplate “direct” and some “circumstantial” evidence (e.g., fingerprints), making it seem stronger than it is, and to magnify the fact as to non-exclusionary non-matches, making them seem weaker than they are. We propose ways to use matches and non-matches more effectively to avoid miscarriages.

- I. INTRODUCTION: THE UNDERUSE OF NON-EXCLUSIONARY NON-MATCHES TO AVOID WRONGFUL CONVICTION
- II. THE AGGREGATE POWER OF NON-EXCLUSIONARY NON-MATCHES ON THE QUESTION OF IDENTITY
 - A. *People v. Adamson* Revised
 - B. *R. v. Adams*
- III. JUDICIAL RESISTANCE TO AGGREGATIVE ANALYSIS OF MATCHES AND NON-MATCHES
 - A. Objections to Probabilistic Proof Generally: *People v. Collins*
 - B. Objections to Aggregative Analysis of Matches and Non-Matches
- IV. STANDARD IDENTITY EVIDENCE AS AGGREGATIONS OF MATCHES AND NON-MATCHES
 - A. The Compatibility of Aggregative Analysis and Criminal Justice
 1. Confessions, Eyewitness Identifications and Fingerprints as Aggregative Evidence
 2. The Courts’ Enthusiastic Embrace of Aggregative Analysis: Inculpatory DNA
 - B. The Adversarial System’s Ability to Domesticize Aggregative Analysis
 - C. Modern Data Mining and the Broad Availability of Frequency Information
 - D. Aggregative Use of Non-Exclusionary Non-Matches Outside of Court
- V. COGNITIVE, STRUCTURAL AND LEGAL IMPEDIMENTS TO USING NON-EXCLUSIONARY NON-MATCHES
 - A. Cognitive Resistance
 1. Heuristic Economization
 2. The Representativeness Bias
 3. The Simulation, Confirmation and Certainty Biases

¹ See James Baldwin, *The Evidence of Things Not Seen* (1986).

² Simon H. Rifkind Professor, Columbia Law School.

³ Columbia Law School, J.D. (expected) 2012.

⁴ Columbia Law School, J.D. (expected) 2013.

⁵ Columbia Law School, J.D. 2011. Thanks to Alexandra Blaszczyk, Leslie Demers, Robert King and Abshir Kore for superb research assistance.

- a. The simulation bias
 - b. The confirmation bias
 - c. The certainty effect
 - 4. The “Uniqueness Fallacy”
 - 5. The “Irrelevance Fallacy”
 - B. Structural Disadvantages
 - 1. Reasons to Doubt the Adversarial Antidote
 - 2. The Biasing Effect of the State’s Monopoly over the Initial Investigation
 - C. Legal Obstacles
 - 1. Rules Distinguishing “Direct” and “Circumstantial” Evidence
 - 2. Rules Regulating Police Investigations
 - 3. Rules Regulating Discovery
 - 4. Rules Limiting Evidence of a Third Party’s Guilt
 - 5. Rules Limiting Statistical Evidence
- VI. SENSIBLE REGULATION OF NON-EXCLUSIONARY NON-MATCHES AS EVIDENCE OF IDENTITY
- A. New Tools to Improve Lay Decision Makers’ Appreciation of Aggregative Analysis
 - B. Management-Based Regulation of Non-Exclusionary Non-Matches as Evidence of Identity
 - 1. Management-Based Regulation
 - 2. Regulation of “Big” Evidence of Identity
 - 3. Regulation of “Small” Non-Matches
- VII. CONCLUSION

I. INTRODUCTION: THE UNDERUSE OF NON-EXCLUSIONARY NON-MATCHES
TO AVOID WRONGFUL CONVICTION

In 1997, a New Orleans jury convicted juveniles Ryan Mathews and Travis Hayes of the robbery-murder of a New Orleans store owner.⁶ There was considerable evidence against the two. Witnesses reported that a black male shot the robbery victim, then leapt into a getaway car through an open passenger-side window and escaped. Police stopped Hayes and Mathews in a vehicle resembling witnesses’ description of the getaway car. Hayes confessed to police that he and Mathews planned the robbery in which he drove the car and Mathews, high on marijuana, entered the store and ran out after several gunshots. A witness identified Mathews as the man he saw running from the store.

Although strong, the evidence was deficient in some respects. Hayes was borderline mentally retarded and confessed after six hours of interrogation. His statements were modestly inconsistent with each other and with known details of the crime. The make and model of the car Hayes was driving when

⁶ See *State v. Hayes*, 806 So. 2d 816, 821 & n.10 (La. Ct. App. 2001), Bob Herbert, *Trapped in the System*, N.Y. Times, Jul. 14, 2003, at A17; Michael Perlstein, *A Forgotten Man: Prosecutors Refuse to Reconsider Inmate’s Case Despite Evidence Supporting his Claim*, New Orleans Times-Picayune, Oct. 10, 2004, at 1.

the two were arrested were not the same as witnesses reported, and the car had a rolled up passenger side window that had been inoperable for as long as anyone could remember. The eyewitness who identified Mathews watched the events through a rear view mirror, and other eyewitnesses could not identify Mathews and described the shooter as 5'6", seven inches shorter than Mathews. DNA on a ski mask left at the crime scene matched neither defendant. None of the deficiencies was inconsistent with guilt, however, and the jury convicted both men. The jury sentenced Hayes to life in prison and Mathews to death.

Shortly after the trial, Louisiana convicted Rondell Love of an unrelated manslaughter and sentenced him to 20 years in the Louisiana State Penitentiary, where Mathews was awaiting execution. Love told other inmates that he had committed the murder for which Matthew was condemned. Mathews got wind of Love's claims and reported them to his lawyers who obtained DNA testing of the ski mask at the scene. The results implicated Love, who at 5'7" and medium build matched the witnesses' descriptions. Mathews and Hayes were eventually exonerated.⁷

In 1983, a jury convicted Carlos DeLuna of the stabbing death of a young female Hispanic clerk at a Diamond-Shamrock gas station in Corpus Christi, Texas.⁸ The main evidence at the trial was a night-time show-up identification of DeLuna by the sole eyewitness to the single-perpetrator crime who had watched the assailant escape on foot. Again, there were discrepancies, including the eyewitness's initial description of the shabbily dressed, mustachioed and bewhiskered "derelict" he saw struggling with the victim, fingerprints at the scene and the blood-soaked scene itself, none of which matched the blood-free white dress shirt, dress pants and shoes worn by the clean-shaven DeLuna when he was arrested shortly after the killing. The crime was captured on a 911 phone call from the store clerk. Police arrested DeLuna

⁷ Elisabeth Salemme, Gallery of the Exonerated: Travis Hayes and Ryan Mathews, Time, June 5, 2007, 2007, at http://www.time.com/time/specials/2007/article/0,28804,1627368_1627366_1627389,00.html (last visited Dec. 1, 2011) (noting that Mathews was exonerated in June 2004 and released in December 2006, following eight years in prison).

⁸ See James S. Liebman, et al., *Los Tocayos Carlos*, ___ COLUM. HUM. RTS. L. REV. ___ (2012) (forthcoming); Steve Mills & Maurice Possley, I Didn't Do It. But I Know Who Did: New Evidence Suggests a 1989 Execution in Texas Was a Case of Mistaken Identity, Chi. Trib., June 25, 2006, at C20; Steve Mills & Maurice Possley, A Phantom, or the Killer?: A Prosecutor Said Carlos Hernandez Didn't Exist. But he Did, and his MO Fit the Crime, Chi. Trib., June 26, 2006, at C1; Steve Mills & Maurice Possley, Did One Man Die for Another Man's Crime?: The Secret That Wasn't, Chi. Trib., June 27, 2006, at C1.

three blocks away, cowering under a pick-up truck. DeLuna testified that he had seen an acquaintance, Carlos Hernandez, wrestling with the clerk inside the store and fled when he heard sirens coming because he had been drinking in violation of his parole conditions. A police officer testified that he scoured police records for a “Carlos Hernandez” matching the eyewitness description of the assailant but found none. The prosecutor dubbed Carlos Hernandez a “phantom” and Carlos DeLuna a “liar” for fabricating him. During DeLuna’s appeals, an affirming court expressed “substantial doubt that Carlos Hernandez even existed.”⁹ DeLuna was quietly executed in 1989.

Seventeen years later, the senior author of this article, and later the *Chicago Tribune*, reinvestigated the case and discovered that Carlos Hernandez existed and was well-known to Corpus Christi law enforcement.¹⁰ Three years before the convenience store stabbing, the detective and assistant district attorney handling that case had considered Carlos Hernandez a prime suspect in the beating and knifing death of another young Hispanic woman.¹¹ Although the detective and prosecutor eventually arrested and tried a different man, he was acquitted after his attorney marshaled evidence that Carlos Hernandez was the culprit. Like DeLuna, Hernandez was 5’8” tall and weighed 160 pounds, as was the man the eyewitness described to police immediately after the crime. In the subsequent reinvestigation, relatives of both Carloses mistook one for the other when shown pictures of the two taken within weeks of each other and of the gas station killing (*Figure 1* below).¹² Shortly after the killing, and for years afterwards, Hernandez told associates that he, not DeLuna, committed the crime.¹³ Eight months before DeLuna was executed, Hernandez stabbed another young Hispanic woman nearly to death with a lock-blade buck knife identical to the one found at the Diamond-Shamrock crime scene.¹⁴ Hernandez had

⁹ DeLuna v. McCotter, No. 86-234, at 8 (S.D. Tex. Jun. 13, 1988).

¹⁰ Liebman et al., supra note 8, at ___.

¹¹ Id. at ___.

¹² Id. at ___. The profile view in the photographs is the same one the eyewitness had as he watched the assailant flee.

¹³ Id. at ___.

¹⁴ Id. at ___.

previously confessed the Diamond-Shamrock stabbing to this woman, and in the midst of attacking her, told her she was going to suffer the same fate because she insisted on dating another man.¹⁵

Hernandez's characteristic modes of dress and grooming at the time of the convenience store and other traits that did not match DeLuna did match the eyewitness's initial description of the assailant.¹⁶

Figure 1



Figure 1. Carlos DeLuna (left) two weeks before the killing for which he was executed. Carlos Hernandez (right) two months after being arrested with a knife behind another convenience store. Hernandez wore a moustache his entire adult life except in the weeks after the Diamond-Shamrock killing.

When considering what went wrong in cases like *Mathews/Hayes* and *DeLuna*, commentators focus on the mishandling of what might be called “big” evidence of identity—evidence that by itself is likely to impress a jury, such as Hayes' confession and the show-up identification of DeLuna.¹⁷ These observers point out that identifications,¹⁸ confessions¹⁹ and “snitch” testimony²⁰ (another classic example

¹⁵ Id. at __.

¹⁶ See *infra* notes 26 and accompanying text (analyzing matching and non-matching evidence in DeLuna). Hernandez died in prison in 1999. Liebman et al., *supra* note 8, at __.

¹⁷ See, e.g., *Arizona v. Fulminante*, 499 U.S. 279, 296 (1991) (“A confession is like no other evidence [and] ‘is probably the most probative and damaging evidence that can be admitted against [a defendant].’” (citation omitted)); *Colorado v. Connelley*, 479 U.S. 157, 182 (1986) (“[A] confession makes the other aspects of a trial superfluous and the real trial, for all practical purposes, occurs when the confession is obtained.”).

¹⁸ See, e.g., Brian L. Cutler & Steven D. Penrod, *Mistaken Identification: The Eyewitness, Psychology, and the Law* 13 (1995) (describing studies suggesting that eyewitness identifications are mistaken over half the time); Brandon L. Garrett, *Convicting the Innocent: Where Criminal Prosecutions Go Wrong* 45-83 (2011) (linking eyewitness misidentifications revealed by DNA exonerations to suggestive police procedures, including stacked lineups and prejudicial remarks by police to witnesses); Sandra Guerra Thompson, *Beyond a Reasonable Doubt? Reconsidering Uncorroborated Eyewitness Identification Testimony*, 41 U.C. Davis L. Rev. 1487 (2008); Kevin Jon Heller, *The Cognitive Psychology of Circumstantial Evidence*, 105 Mich. L. Rev. 241, 248-49, 252-54 (2006) (documenting high error rates in eyewitness identifications); Katherine R. Kruse, *Instituting Innocence Reform: Wisconsin's New Governance Experiment*, 2006 Wis. L. Rev. 645, 652-55 (summarizing the literature); Gary L. Wells & Elizabeth A. Olson, *The Other-Race Effect on Eyewitness Identifications; What Do We Do About It?*, 7 Psychol., Pub. Pol'y & L. 230 (2001).

of influential identity evidence) are less reliable than jurors think. To cut such evidence down to size, these critics advocate expert testimony about the foibles of eyewitness identifications and confessions,²¹ cautionary instructions about informant testimony²² or exclusion of evidence unless it was collected through state-of-the-art techniques such as double-blind line-ups and videotaped confessions.²³

Exoneration cases such as Hayes/Mathews and reinvestigation cases such as DeLuna reveal a second type of indicative evidence, which commentators typically ignore. In almost all these cases, police arrested, prosecutors tried and jurors convicted the defendant despite multiple “non-exclusionary non-matches” between the defendant and potentially evidential traces from the crime scene that later were matched to the “real killer.” “Non-exclusionary non-matches” arise when the suspect or defendant demonstrably was not the source of bits of potential evidence associated with a crime that *might* have been left by the perpetrator but might also have appeared for reasons having nothing to do with the crime. Police, prosecutors and jurors no doubt acted against the defendant, despite the non-matches, precisely because the non-matches were “non-exclusionary”: they had explanations consistent with the defendant’s guilt, and thus did not *exclude* him as the guilty party or prove his innocence.²⁴

¹⁹ See, e.g., Garrett, *supra* note 18, at 22-33 (describing interrogation procedures eliciting false confessions); Heller, *supra* note 18, at 248-49, 254 (2006) (summarizing research indicating that jurors give excessive weight to confessions); Richard A. Leo et al., Bringing Reliability Back In: False Confessions and Legal Safeguards in the Twenty-First Century, 2006 Wis. L. Rev. 479 (2006).

²⁰ See, e.g., Garrett, *supra* note 18, at 123-39; Alexandra Natapoff, Snitching: Criminal Informants and the Erosion of American Justice (2009); Stephen S. Trott, Words of Warning for Prosecutors Using Criminals as Witnesses, 47 Hastings L.J. 1381, 1394 (1996); Ellen Yaroshesky, Cooperation with Federal Prosecutors: Experiences of Truth Telling and Embellishment, 68 Fordham L. Rev. 917, 953 (1999).

²¹ See, e.g., James M. Doyle, Learning From Error in American Criminal Justice, 100 J. Crim. L. & Criminol. 109, 113 (2010) (advocating expert testimony about problems with eyewitness identifications); Cindy J. O'Hagan, When Seeing is Not Believing: The Case for Eyewitness Expert Testimony, 81 Geo. L.J. 741, 744 (1993) (advocating expert testimony on difficult memorization process required in making identifications and effect of light, movement, duration of encounter and weapons). See also *People v. Santiago*, 17 N.Y.3d 661, 668-74 (N.Y. 2011) (overturning conviction because trial court excluded expert testimony on risks accompanying eyewitness identifications that are cross-ethnic, expressed with confidence and based on observations made at gunpoint).

²² See, e.g., Aaron M. Clemens, Removing the Market for Lying Snitches: Reforms to Prevent Unjust Convictions, 23 Quinnipiac L. Rev. 151, 235 (2004) (proposing polygraph examinations and other ways to deter false statements by jailhouse informants).

²³ See, e.g., Alan Gershel, A Review of the Law in Jurisdictions Requiring Electronic Recording of Custodial Interrogations, 16 Rich. J. L. & Tech. 9 (2010) (advocating videotaping of interrogations); Kruse, *supra* note 18, at 660 & n.71, 661-62 (discussing videotaping of interrogations and sequential, double-blind eyewitness line-ups); John Schwartz, Changes to Police Lineup Procedures Cut Eyewitness Mistakes, Study Says, N.Y. Times, Sept. 19, 2011, at A13 (describing evidence that sequential, double-blind lineups “catch fewer innocent suspects”); Thomas P. Sullivan, Police Experiences with Recording Custodial Interrogations, 88 Judicature 132, 132-33 (2004).

²⁴ *Excusatory* non-matches exonerate the defendant when the culprit must have left a trace that the defendant could not have left. Examples are non-matching single-source DNA found in semen in the vagina of a rape victim immediately after the crime,

Examples include the non-match in the *Hayes/Mathews* case between the make and model of the car witnesses linked to the killing and the car Hayes was driving, between the open passenger-side window of the car through which the robber was said to have jumped and the fact that the passenger window on Hayes's car had been stuck in a rolled-up position for years, and between the 5'6" assailant witnesses described and Mathews 6'1" frame.²⁵ Examples from the *DeLuna* case are collected in *Figure 2*. Of thirty traces associated with the crime, three match neither Carlos DeLuna (executed for the crime), nor Carlos Hernandez (who told associates he committed the crime). Eight more were never tested as to either man, due to police inaction—for example, large amounts of blood found at the crime scene and evidence visible in the investigating officers' photographs that they never noticed, including bloody shoe prints, a wad of chewed gum spat onto the floor and a clump of hair. Only seven of the thirty traces match DeLuna, all seven of which also match Hernandez: height, weight, ethnicity, sex, hair color, hair style and cigarette brand. Of the twelve remaining *non-matches* as to DeLuna, two were never tested as to Hernandez (now deceased), and the remaining ten match Hernandez—including age, clothing, moustache and weapon of choice. Although the jury convicted DeLuna based in part on the seven matches, it never knew that seventeen or more traits matched Hernandez, including all seven that matched DeLuna.

or non-matching fingerprints encased in the victim's blood on the murder weapon. Even these matches are not absolutely exclusionary, given the possibility of a frame-up or laboratory mistake.

²⁵ See Lisa L. Smith et al., *Understanding Juror Perceptions of Forensic Evidence*, 56 *J. Forensic Sci.* 409, 409 (2011) ("Current forensic science techniques are capable of analyzing a wide range of materials (e.g., glass, fibers, paint, gun-shot residue) that can be used to establish a connection between a source and a criminal act or crime scene [and have greatly] decreased the quantity of trace material required to conduct useful comparative analyses.").

Figure 2²⁶

	Eyewitness description of killer	Traces found by police at crime scene	Traces not found by police but visible in their photos or seen by others	Man nearby seen fleeing or acting suspiciously	Tot.
Matches DeLuna, not Hernandez					0
Matches DeLuna and Hernandez	1. 5'8" 2. 160 lbs 3. Hispanic 4. Male 5. Black hair 6. Wavy hair	14. Winston cigarettes			7
Matches Hernandez, not DeLuna	7. Mid/late 20s 8. Moustache 9. Unshaven 10. "Derelict" (shabby, unkempt clothes) 11. Red plaid flannel shirt or jacket 12. White or light grey sweatshirt 13. Blue jeans	15. Lock-blade buck knife 16. Maroon button		28. Man seen earlier lurking outside gas station at McArdle & Kostoryz	10
Does not match DeLuna; untested as to Hernandez		17. Cash drawer short \$20-\$60	23. Clump of hair on floor		2
Untested as to both		18. Blood pool, smears, spatter 19. Cigarette fragment	24. Bloody palm prints 25. Bloody shoe prints 26. Wad gum 27. Cement chunks	29. Man fleeing north on Dodd St. 30. Man fleeing from Dodd & McArdle to McArdle & Kostoryz	8
Does not match either		20. Fingerprint - phone 21. Fingerprint - door 22. Fingerprint - beer can			3

It is not surprising that police, prosecutors and jurors in the *Hayes/Mathews* and *DeLuna* cases dismissed these "small" non-matches—small because that they were easily explained away on grounds unrelated to the defendant's guilt or innocence—and focused instead on confessions, eyewitness identifications and other "big" evidence of identity. No one begrudges a prospector for gold who ignores tiny flecks of the stuff in a stream bed while seeking the mother lode up the canyon side. Compared to the tedious task of accumulating enough flecks to add up to a real stake, the prospect of striking it rich with the discovery of a single large vein is hard to pass up. This article argues, however, that systematically

²⁶ Source: Liebman, et al., supra note 8, at ___ & Figure ___.

aggregating more of the small flecks, and putting less faith in the deceptive allure of big lodes of evidence, could generate more evidence of identity and a decrease in wrongful conviction.

In Part II we use both hypothetical and actual examples to illustrate how systematically aggregative analysis of non-matching bits of evidence might improve the accuracy of answers police, prosecutors and jurors reach on the “whodunit?” question. The English courts’ rejection of aggregative analysis in the actual case discussed in Part II prompts our examination in Part III of the reasons courts give for resisting efforts to quantify the conjoint effect of small bits of non-dispositive evidence the identity question. The classic case is *People v. Collins*,²⁷ which overturned a California robbery conviction premised in part on a prosecutor’s effort to quantify the overall effect of several individually non-dispositive *matches* between the two defendants in the case and what was known about the interracial couple that committed the robbery. Although decided over forty years ago and easily distinguished, given the prosecutor’s flawed methodology and barely concealed racial ulterior motive, the case continues to provide the intellectual basis for resisting probabilistic proof generally.

Part IV responds to the *Collins* critique by showing that the courts themselves have rejected it in actual practice through their enthusiastic embrace of inculpatory DNA evidence. With DNA evidence as our prime example, and confession, eyewitness identification and fingerprint evidence as supporting cases, we elucidate the often acknowledged fact that all evidence is probabilistic²⁸ with a demonstration of a less obvious fact: The high probabilities associated with these and other supposedly “unique” traits matching suspect to culprit are in fact the aggregate result of multiple matches of non-unique, often very common traits. Other things equal, the more non-unique matching traits subsumed by any given DNA sample, confession, eyewitness identification, fingerprint or other piece of “big evidence,” the more probative weight the evidence has. DNA evidence reveals, as well, that there is no reason in principle to

²⁷ *People v. Collins*, 438 P.2d 33 (Cal. 1968).

²⁸ See, e.g., Richard A. Posner, *An Economic Approach to the Law of Evidence*, 51 *Stan. L. Rev.* 1477, 1508 (1999) (“It is now generally recognized, even by the judiciary, that since all evidence is probabilistic—there are no metaphysical certainties—evidence should not be excluded merely because its accuracy can be expressed in explicitly probabilistic terms.”); *infra* notes 86 and accompanying text.

shield even lay jurors from the formal quantification of the aggregate effect of non-unique bits of evidence into a stated probability that the defendant was responsible for the existence of evidence associated with a crime. DNA thus gives the lie to *Collins*' most far-reaching claim, that formal analysis of the aggregate effect of non-unique matching traits is anathema to accepted modes of judicial fact finding. As Part IV also develops, the adversary system's swift and dramatic improvement of forensic DNA analysis, along with the burgeoning capacity of data-mining techniques to reveal the frequency of millions of traits in relevant populations and environments, neutralize *Collins*' lesser, more technical reasons for rejecting formal aggregative analysis in criminal trials. What is true for trials is even truer for investigations and prosecutorial decision making, where our commitments to adversarial judicial proceedings and jury decision making are not limiting factors.

There are, however, other obstacles to using aggregative analysis of matches and non-matches in criminal investigations, trials and appeals. Part V addresses three categories of barriers —cognitive, structural and legal. Focusing mainly on the use of non-exclusionary non-matches to decrease the likelihood of arresting, charging and convicting the innocent, Part V shows how heuristic economization interacts with the structure of trials to reinforce legal resistance to the aggregative use of “small” evidence and discourage use of the adversarial system to discipline and improve aggregative analysis. In Part VI, we chart two intersecting paths around these obstacles. One is provided by emerging tools designed to make aggregative analysis more intuitively accessible to investigators and jurors. The other is the use of management-based regulation, along with the adversarial system, to discipline systematic steps to pan for small but cumulatively powerful flecks of evidence of identity in criminal investigations and trials.

II. THE AGGREGATE POWER OF NON-EXCLUSIONARY NON-MATCHES ON THE QUESTION OF IDENTITY

The *Hayes/Mathews* and *DeLuna* cases suggest that fuller attention to non-exclusionary non-matches might avoid miscarriages of justice. This Part uses two examples—one stylized, the other from

an actual case—to show how systematically aggregating the effect of multiple non-matches, none of which is very probative by itself, can generate more accurate results.

A. *People v. Adamson Revised*

Although Anglo-American courts resist such analysis, Bayes' Theorem may be used to demonstrate deductively the power of non-exclusionary non-matches to distinguish probably guilty from probably innocent defendants whom intuitive decision makers would likely treat as equally guilty.²⁹

To show how, we use a hypothetical example based on a familiar case, *People v. Adamson*.³⁰ In *Adamson*, police investigated a murder of an elderly white woman in Los Angeles. The victim was found in her home without any stockings on, but the bottom half of a pair of women's stockings was found nearby. The top halves of the stockings were missing. Suspicion came to rest on Dewey Adamson, a black middle-aged man. When police located him and searched his house, they found one cut-off stocking top on his dresser and two more in a drawer. Although the stocking tops in Adamson's possession did not match the bottoms found at the crime scene, evidence of both was admitted against Adamson at trial. On appeal, the California and United States Supreme Courts rejected Adamson's evidentiary and due process claims that the stocking evidence should have been excluded as more prejudicial than probative.³¹

The case is a favorite of evidence teachers because it illustrates the judgment calls and cultural biases that can afflict the assessment of probative weight and prejudice.³² When the case arose in the 1940s, evidently the only source of both probative weight and prejudice that the white lawyers and judges in the case could identify was an inference of sexual perversion—a man's interest in women's stocking

²⁹ See 1 David L. Faigman et al., *Modern Scientific Evidence: The Law and Science of Expert Testimony* 236-38 (2005) (describing deductive nature of Bayesian logic); Athanasios Papoulis, *Probability, Random Variables, and Stochastic Processes*, 38-39, 78-81, 112-14 (1984) (deriving Bayes' Theorem mathematically).

³⁰ 165 P.2d 3 (Cal. 1946), *aff'd*, 332 U.S. 46 (1947). The United States Supreme Court's decision in *Adamson v. California*, 332 U.S. 46, 58 (U.S. 1947), is best known for the debate it engendered over whether the Fourteenth Amendment Due Process Clause incorporates the Fifth Amendment. Compare *id.* at 59-67 (Frankfurter, J., concurring) (resisting incorporation) with *id.* at 68-92 (Black, J., dissenting) (advocating incorporation).

³¹ See *Adamson v. California*, 332 U.S. at 59 (concluding that the California courts' holding that the stocking tops were admissible as evidence because this "'interest in women's stocking tops is a circumstance that tends to identify defendant'" as the perpetrator did not violate the Constitution (quoting *People v. Adamson*, 165 P.2d at 6)).

³² See, e.g. Jack B. Weinstein et al., *Evidence Cases and Materials* 6 (9th ed. 1997); Eric D. Green et al. *Problems, Cases, and Materials on Evidence* 69 (3d ed. 2000).

tops. The all-white jury in the case³³ also may have seen only that type of “match” between the culprit and Adamson. The problem, of course, is that within the African-American community at the time,³⁴ stocking tops were frequently used by black men in the process of “conking” or using chemicals to straighten their hair.

The case also illustrates the mathematical definitions of relevance, probative weight and prejudice that Richard Lempert famously introduced.³⁵ Under Lempert’s definition, evidence is relevant as long as two probabilities are different—the probability that the evidence would exist if the defendant is the perpetrator and the probability that the evidence would exist if the defendant is not the perpetrator. Stated mathematically, evidence is irrelevant if $P(E/G) / P(E/Not-G) = 1$. Probative weight then is the absolute value of the difference between the numerator and denominator values divided by the numerator. Additionally, one form of prejudice—“misestimation”—is defined by the difference between that sum, and the sum the jury is likely to arrive at intuitively, for example, by miscalculating the denominator value (the probability the evidence would exist though the defendant is not guilty) because, for example, the jury doesn’t know that African-American men often use women’s stocking tops as a hair care implement.³⁶

We modify the *Adamson* example for another purpose: to illustrate the use of Bayes’ Theorem to calculate the aggregate value of non-exclusionary non-matches and demonstrate the existence of reasonable doubt that otherwise might escape the decision maker’s attention. We begin with an insight of Professors Eric Green and Charles Nesson.³⁷ They used *Adamson* to identify an evidentiary mistake that—although they didn’t describe it this way—undermines the accuracy of intuitive estimates of the numerator probability in the likelihood-ratio calculation of relevance and probative weight. Building on

³³ *People v. Adamson*, 210 P.2d 13, 15 (Cal. 1949) (denying Adamson’s claim that the all-white jury that convicted him violated his equal protection and due process rights).

³⁴ See Malcolm X & Alex Haley, *The Autobiography of Malcolm X* 61-62 (1965).

³⁵ Richard O. Lempert, *Modeling Relevance*, 75 Mich. L. Rev. 1027 (1975).

³⁶ Magnifying the potential for prejudice is the concern that, when the numerator value is high (the probability that if the defendant in a case in which the perpetrator ran off with the victim’s stocking tops is *guilty*, we would find that he has an interest in women’s stocking tops), jurors may forget to ask the denominator question (how often *innocent* defendants may have an interest in women’s stocking tops). See *infra* notes 234-259 and accompanying text.

³⁷ Eric D. Green & Charles R. Nesson, *Problems, Cases, and Materials on Evidence* 70-71 (2d ed. 1994).

their insight, we suppose that the case arose in 2012, not 1946, and that Adamson is white, not black. Police arrive at the victim's home, where she lived alone, five minutes after a neighbor reports hearing her scream. The police immediately spot several attributes of the crime scene that provide clues to the identity of the culprit: (1) missing stocking tops; (2) the victim's wounds, which suggest that the powerful fatal blows were delivered left-handed; (3) a description—"early forties, balding, 5'3" tall"—by a male neighbor, who was one of two people who saw a man standing outside the victim's home as each passed by within minutes of the time a third neighbor heard a scream; and (4) a shopper's complaint to a police officer on foot a block away, soon after the scream was heard, that "this white guy with green eyes just ran hell-bent down the street, nearly knocking me over as I came out of the supermarket" and sped off in the same direction, away from the victim's home. Shortly after that, police get an anonymous tip that Adamson committed the crime and go to his nearby apartment to talk to him. They bring the two neighbor-witnesses with them. Adamson steps outside his apartment where he lives alone, giving the witnesses a good look at him. The male neighbor says he can't say whether Adamson is the man he saw outside the victim's house. The female neighbor identifies Adamson as the man she saw there a few minutes after the male neighbor passed by and two minutes before the scream was heard.

The police arrest Adamson. They search his home and find several sets of women's stocking tops in his possession, none matching the stocking bottoms found near the victim. Adamson is 29-years-old, 5'8", right-handed, with a full head of hair and brown eyes.

At trial, the state calls the female neighbor to testify that she saw a man outside the victim's home and accompanied police to Adamson's house and identified him there, and that she is sure he is the man she saw at the victim's home. Thereafter, the prosecution introduces evidence of the missing stocking tops at the crime scene and the sets of women's stocking tops found in Adamson's home. A detective testifies that she initially concluded that the fatal blows to the victim were from the assailant's left hand, but now backs off of that conclusion, noting that a right-handed assailant could have struck the fatal blows with his left hand. The defense then presents evidence establishing Adamson's height, weight, eye

color and right-handedness and calls both the male neighbor to describe the 5'3", balding forty-year-old man he saw near the crime scene, and the shopper who was nearly knocked over by a green-eyed sprinter headed away from the crime scene.

In closing argument, the prosecutor dismisses the men those two witnesses described as different from the man the female neighbor saw at the victim's house just before the crime and identified as Adamson. Based on that identification and the subsequent discovery that Adamson shares the culprit's interest in women's stocking tops, the prosecutor urges the jury to convict him.

As Professors Green and Nesson suggested,³⁸ there are multiple dangers in this situation. First, the jury may underestimate the *denominator* probability that innocent men have a use for women's stocking tops. Second, the inference of perversion from the stocking-top evidence may prejudice the jury against the defendant irrespective of his connection to the crime.

Less obviously, however, the jury also may overestimate the *numerator* probability—that a suspect arrested for a crime in which the victim's stocking tops disappeared would be found to have stocking fragments. The proper estimation of the numerator in fact is how probable it is that the suspect would match that stocking tops clue but *not* match the other clues—for example, the left-handed blows the detective initially associated with the killing; the male neighbor's description of a man at the victim's house as a short, balding forty-year-old; and the shopper's encounter with a green-eyed man running hell-bent away from the crime scene. But, as Green and Nesson pointed out,³⁹ once the defendant is arrested based on evidence that *does* match him—once the case turns from a “Whodunit?” to a “Did Adamson do it?”—we can expect participants in the case to limit the numerator question to the matching stocking-top evidence and ignore the non-matching handedness, height, hair, age and eye-color clues. The same would be true if, for example, (1) police had found a partially smoked cigarette on the floor just inside the home

³⁸ Id. at 70-71 (listing “risk[s] of overvaluing” trace evidence matching the defendant: (1) at trial, where only a single suspect is in view, matching evidence looms larger than it more appropriately does during investigations where the possibility of multiple suspects, many of whom may match the trait, is front of mind; (2) too little attention is paid to evidence at the scene that does not match a known suspect; and (3) leaving it “to the defense to fill out the context to offset any tendencies of the evidence to mislead” exposes defendants “to risks of inadequate defense representation”).

³⁹ See id.

of the victim (a nonsmoker), (2) the female neighbor had told police that the man at the victim's house wore a brown sweatshirt, and yet (3) police found no brown sweatshirt in Adamson's possession or evidence that he smoked.

We can now apply Bayes' Theorem to this embellished hypothetical to demonstrate that the attention the trial participants do or (predictably) do not give to the non-matching evidence could be the difference between an accurate, reasonable-doubt acquittal and a false conviction. Suppose that after hearing the female neighbor's testimony—and taking due account of the fact that the man she saw might not have been the killer, and that her albeit confident and unshakeable identification might be mistaken—the jury concludes there is an 85 percent chance Adamson was the killer. After hearing additional evidence that the victim's stocking tops were missing and Adamson had non-matching stocking tops in his possession, the jury raises the probability that Adamson is guilty to 98 percent and is prepared to convict. The jury (predictably) is not much moved by defense evidence and argument emphasizing the non-exclusionary non-matches, and thereafter lowers the probability of guilt only to 97 percent.

Finally, suppose the trial judge takes two highly unusual steps. First, she gets special verdicts from the jury with the probabilities noted above. Second, at a hearing on Adamson's motion for relief from a judgment of conviction, she lets a defense expert present evidence, based on census data and information mined from nearby surveillance cameras and retail sales, that 11 percent of adult males in the area are left-handed or ambidextrous; 5 percent own brown sweatshirts; and 10 percent smoke cigarettes. Based on this data, and on the 98 percent probability of guilt that the jury estimated after hearing the prosecution's case, the expert calculates the effect on the jury's estimate of a proper evaluation of the five non-matching traits—the left-handedness clue; the male neighbor's description of a short, balding forty-year-old at the victim's house; the shopper's encounter with a green-eyed sprinter; the brown sweatshirt; and the cigarette.

The expert begins with Bayes' Theorem, which demonstrates mathematically that the probability of an event—here Adamson's guilt—after each new bit of evidence is the *prior odds* of the event times

the *probative value* of the new bit of evidence. Probative value is measured using the *likelihood ratio* introduced above: the probability that the new piece of evidence would be present if the defendant is guilty ($P(E/G)$) divided by the probability that the same evidence would appear if the defendant is not guilty ($P(E/\text{not } G)$):⁴⁰

(prior odds of guilt) x (likelihood ratio associated with new evidence) = subsequent odds of guilt

Or: *(prior odds of guilt) x ($P(E/G) / P(E/\text{not } G)$) = subsequent odds of guilt*

The expert treats the prior odds of guilt as 98 to 2, given the 98 percent probability of guilt the jury found after hearing the state's case. The expert then analyzes the effect of each non-match, starting with detective's initial belief that the killer administered the fatal blows with a left fist. Laying aside credibility issues for now, the expert estimates the likelihood-ratio numerator—the chance that the assailant would have delivered such powerful blows left-handed if he, like Adamson, were right-handed—as 5 percent. The expert then estimates the denominator probability—that the evidence would be present if someone other than Adamson committed the crime—as 15.45 percent. This is the rate of non-right-handers in the population (11 percent), plus the rate of right-handers (89 percent) times the five-percent possibility that a right-hander used his left hand to beat the victim ($.05 \times 89 = 4.45$).

As for the male witness's description of a short, balding forty-year-old outside the victim's home, the expert posits a 75 percent probability that, assuming Adamson is guilty, someone would have seen a man outside the victim's house just before the crime and recalled him as 5'3", balding and forty-ish (though Adamson is 5'8", hirsute and 29 years old). Assuming that a witness's recollection of the features of a man seen recently are somewhat more likely to be accurate than not, the expert rates the denominator probability—that the "short, bald, fortyish" description would have been made if someone besides Adamson were guilty—as slightly higher, 80 percent.⁴¹ The expert likewise rates the probability of a

⁴⁰ For additional discussion of Bayes' Theorem, see *infra* notes 217-218, 238 and accompanying text. On the likelihood ratio as a measure of probative value, see *supra* notes 35-36 and accompanying text.

⁴¹ If there is a 75 percent chance, assuming Adamson's guilt, that a witness would have seen someone—Adamson or another person—outside the victim's home and described the person as short, bald and forty-ish, though Adamson is none of those things, then the chances must be higher that the same thing would have happened if someone else were guilty. The two possibilities are

woman seeing someone running from the crime scene who is not Adamson (given his non-matching eye color) as lower if Adamson is guilty (the numerator) than if he is not guilty (the denominator), because the latter but not the former probability includes the possibility that the sprinter was the fleeing assailant. The expert estimates the numerator as 10 percent and the denominator as 20 percent.

Next, given that neither the victim nor Adamson smokes, the numerator probability of finding a cigarette in the victim’s house if Adamson is guilty is fairly low, say 20 percent. But if Adamson is not guilty, the probability is higher that the killer left a cigarette behind at the scene, say 30 percent. Finally, reasoning that the killer likely discarded his tell-tale outer clothing while escaping, the expert estimates a 70 percent numerator probability that, assuming Adamson is guilty, police would have found no brown sweatshirt around him when he was arrested, though a brown-sweat-shirted man was seen at the victim’s house just before she screamed. Based on information mined from surveillance cameras and retail sales revealing that only 5 percent of men in the area own brown sweatshirts, the denominator probability, that the police would find no brown sweatshirt on or around an innocent Adamson is 95 percent.

The expert then uses Bayes’ Theorem to calculate the effect on the prior odds of each non-match, starting with the left-handed blows, which reduce the odds of guilt from the jury’s 98-to-2 (98 percent) to 15.8-to-1, or 94 percent:

$$(98/2)(5/15.5) = 490/31 = 15.8:1 \rightarrow 15.8/(15.8 + 1) = 94.1\%$$

Next, the male witness’s description of the short bald man reduces the 15.8-to-1 odds of Adamson’s guilt to 14.1-to-1, or 93.7 percent:

$$(98/2)(5/15.5)(75/80) = (36,750/2,480) = 14.8:1 \rightarrow 93.7\%$$

In like fashion, the shopper’s encounter with a green-eyed sprinter reduces the probability to 88.1 percent; the cigarette reduces it to 83.2 percent, and the sweatshirt evidence reduces it further to 78.5 percent.⁴²

the same, except that the former includes the chance that the man seen was the killer (Adamson) who was badly misdescribed, and the latter includes the somewhat greater possibility that the man seen was the killer who was accurately described.

⁴² The effect of the five non-matches is as follows, depending upon the starting odds after the state presented its case:

	Percent Probability of Guilt		
Prior probability (ID + stocking tops)	98	99	97

When considered in isolation, the expert notes, none of the non-matches is very probative. In the aggregate, however, the evidence may be powerful enough to reduce the probability of innocence from a negligible 2 percent to more than a one-in-five chance, which would give most observers a reasonable doubt. The expert acknowledges that little weight should be placed on her precise estimates, including because they ignore the possibility that the jury doubted the veracity of one or another witness. The expert's point, however, is not that her estimates and calculations are exact—a jury would be free to substitute others—but only that the non-matches make a lot more difference in the aggregate, than separately. If, therefore, our hypothesis is correct that the jury, unaided, is likely to consider and dismiss each “small” non-match piecemeal—and especially if the jury never hears about some of them and downgrades the credibility of others by heuristic mistake, given excessive confidence in the story told by the prosecutors’ “big” evidence—there is reason to doubt the jury’s guilt determination. Part V provides support for each of these “if” statements.

B. *R. v. Adams*

Our stylized *Adamson* example illustrates how aggregative analysis of non-exclusionary non-matches may reveal that a seemingly straightforward case for a guilty verdict actually merits a reasonable doubt favoring the defendant. A recent English case, *R. v. Adams*, demonstrated the same potential in practice. Although, channeling *People v. Collins*, the British courts ultimately rejected aggregative analysis, an examination of *Adams* suggests what a number of British applied statisticians have said about the result, that the cause of truth was ill-served.⁴³

+ Left-handed blows	94.1	97.0	91.3
+ Short bald man at scene	93.7	96.8	90.7
+ Runner passing shopper	88.1	93.8	83.1
+ Cigarette	83.2	90.9	76.6
+ Brown sweatshirt	78.5	88.1	70.7

⁴³ See, e.g., Peter Donnelly, *Appealing Statistics*, 47 *Med. Sci. Law* 14, 14 (2007); L. A. Foreman et al., *Interpreting DNA Evidence: A Review*, 71 *Intern'l Stat. Rev.* 473 (2003); J. P. Henderson, *The Use of DNA Statistics in Criminal Trials*, 128 *Forensic Sci. Intern'l* 183 (2002); Michael Lynch et al., *Truth Machine: The Contentious History of DNA Fingerprinting* 218, 219 (2008).

Adams is a “cold hit” DNA case. Police recovered DNA from the crime scene but had no immediate suspects to try to match to the genetic profile. Instead, they trawled a law enforcement database containing the genetic profiles of millions of previously arrested and other individuals to see if any matched the sample from the crime scene.⁴⁴

A DNA cold hit is much weaker evidence of guilt than a DNA match between genetic material found at a crime scene and a sample taken from a suspect who was arrested for reasons apart from his genetic make-up. In the latter case—the one usually associated with DNA matches—the question posed to police or a jury is how likely it is that the match is coincidental. How likely is it that a single individual chosen at random from the population would have the same rare but not unique collection of genetic traits as those found at the crime scene? As can be demonstrated using the most straightforward of mathematical analyses, the so-called “multiplication rule,” which multiplies the probabilities of each of the markers to get the aggregate probability associated with the entire set—that number often is very small.⁴⁵ If a vat has a billion balls, each with a number on it that appears, say, only 133 times in the vat, the chances, on a *single random draw*, of selecting a ball with any pre-selected number is very small—in this example, about one in 7.5 million.

By contrast, in a cold hit case, the defendant is arrested precisely because of his genetic make-up, so the question that interests police and the jury is different: How likely is it that there are two or more people in the pool of all possible suspects—all people in the world capable of committing the crime—with that same collection of genetic markers? Analogously, if we scooped out a load of X balls from our vat of one billion, with X being the number of potential suspects, how likely is it that two or more of those balls would have the same number on it? Even if there is only a one in 7.5 million chance that any given individual has a particular collection of non-unique genetic traits, if the notional suspect pool is

⁴⁴ See David J. Balding & Peter Donnelly, Evaluating DNA Profile Evidence When the Suspect Is Identified Through a Database Search, 41 J. Forensic Sci. 599 (1996); Andrea Roth, Safety in Numbers? Deciding When DNA Alone Is Enough to Convict, 85 N.Y.U. L. Rev. 1130, 1140 (2010).

⁴⁵ See *infra* notes 99-102 and accompanying text.

larger than 7.5 million (and, indeed, even if it is smaller), there is a substantial probability that there are at least two matching suspects.⁴⁶

The rape victim in *R. v. Adams* described her assailant as a white, clean-shaven 20-25 year old, with a local accent and a face so distinctive she was sure she would never forget it.⁴⁷ After the police arrested 37-year-old Denis Adams based on a cold DNA hit, the victim could not pick him out of a lineup and said he was not her attacker.⁴⁸ The prosecution nonetheless went to trial based on the “cold” match of Adams’ nine genetic markers and the same nine found on a vaginal swab taken the night of the rape. Those nine markers, the prosecution told the jury, are present in only 1 in 200 million men.⁴⁹ Adams presented the victim’s statement denying that Adams was her attacker, an alibi corroborated by his girlfriend, and evidence that he had a brother whose DNA might also match but was never tested. At trial, the prosecution’s experts admitted that the chances that Adams and his brother had the same collection of genetic makers—1 in 220—were close to 100,000 times greater than the prosecution’s 1 in 200 million figure.⁵⁰ Defense experts testified that the “1 in 200 million” figure was itself one hundred times too low, due to sampling errors and use of sample sizes too small to provide confidence in prosecution’s estimate.⁵¹

Adams and his attorneys were rightly concerned that the “1 in 200 million” probability based on nine matching markers would convince the jury to dismiss as mistaken the victim’s belief that Adams wasn’t her rapist and reject as perjured the alibi testimony of Adams’ girlfriend. Defense experts offered two responses. First, they countered with their own estimates of 1 in 2 million (assuming sampling

⁴⁶ See *People v. Collins*, 438 P.2d 33, 42-43 (Cal. 1968); David H. Kaye, Rounding Up the Usual Suspects: A Legal and Logical Analysis of DNA Database Trawls, 87 N. Car. L. Rev. 425, 458-59 (2009); See David H. Kaye, *People v. Nelson: A Tale of Two Statistics*, 7 Law, Prob. & Risk 249, 251 (2008); Roth, *supra* note 44, at 1143 nn.58, 59 (describing cases in which police investigated suspects who almost certainly did not commit the crime following a cold hit); Richard Willing, Mismatch Calls DNA Tests into Question, USA Today (Feb. 8, 2000).

⁴⁷ See *R. v Adams*, 2 Crim. App. R. 467, 468 (1996).

⁴⁸ See Donnelly, *supra* note 43, at 14.

⁴⁹ *R v Adams*, 2 Crim. App. R. 467, 468 (1996) (“The prosecution case rested entirely upon expert evidence in relation to the DNA sample which was challenged by the defence.”); see Lynch et al., *supra* note 43, at 195 (citing *Adams* as an example of a DNA “cold hit”).

⁵⁰ *R v Adams*, 1 Crim. App. R. 377, 379 (1997).

⁵¹ *Id.* (discussing defense claims that sample size was too small and that the prosecution expert improperly used a pen to draw in a DNA band); *R v Adams*, 2 Crim. App. R. 467, 470 (1996) (discussing defense estimate of the incidence of the group of nine genetic markers in the population as 1 in 2 million).

problems were corrected) and 1 in 220 (given the possibility that Adams' brother was the culprit).⁵² Second, an expert used Bayes' Theorem to show that even fairly modest chances that the victim and girlfriend were telling the truth could substantially diminish the probability of guilt from 55-to-1, or 98 percent (derived from the 1-in-220 estimate) to 2:1 or only 67 percent.⁵³ In essence the defense expert used Bayesian analysis as a mechanism for diminishing the effect of multiplying the probabilities of the nine non-unique genetic-marker matches by extending the equation to capture the contrary effect of the two non-exclusionary non-matches: the victim's and girlfriend's (alibi) testimony. Disregarding the expert testimony, the jury convicted Adams.

On appeal, the court reversed, finding that the trial judge had not properly instructed the jury on how to use Bayesian analysis.⁵⁴ On remand, the trial judge walked the jury more carefully through the expert's analysis and gave the jurors calculators to use in their deliberations.⁵⁵ The jury convicted again. On a second appeal, again alleging that the jury was improperly instructed, the Criminal Division of the Court of Appeal again affirmed. This time, however, the court banned the use of Bayes' Theorem in future criminal trials. Jurors, it noted, should "evaluate evidence and reach a conclusion not by means of a formula, mathematical or otherwise, but by the joint application of their individual common sense and knowledge of the world to the evidence before them."⁵⁶ English jurors had never before used Bayes' Theorem in their deliberations, the court noted, and allowing them to do so would deflect them from their proper task.⁵⁷

Ironically, this new canonical statement of Anglo-American courts' reluctance to use aggregative analysis to add rigor to a jury's consideration of non-exclusionary non-matches arose in a case in which the *only*, albeit seemingly powerful, evidence of guilt was the statistically aggregated effect of a series of *matches* of non-unique genetic traits. Absent aggregative analysis—absent the use of the very

⁵² See *supra* note 51.

⁵³ *R v Adams*, 2 Crim. App. R. 467, 477 (1996).

⁵⁴ *Id.* at 482.

⁵⁵ See Donnelly, *supra* note 43, at 16.

⁵⁶ *R v Adams*, 1 Crim. App. R. 377, 384 (1997) (quoting *R. v. Adams*, 2 Crim. App. R. 467, 482 (1996)).

⁵⁷ *Id.* at ___.

“mathematical formulas” to inform jury deliberation that the court rejected in the process of upholding Adams’ conviction—there was no basis to convict. We reflect further on this irony in Part IV, after delving more deeply into Anglo-American judicial resistance to aggregative analysis in Part III.

III. JUDICIAL RESISTANCE TO AGGREGATIVE ANALYSIS OF MATCHES AND NON-MATCHES

A. *Objections to Probabilistic Proof Generally: People v. Collins*

The classic American judicial decision rejecting aggregative analysis in criminal cases is *People v. Collins*, announced in 1968⁵⁸ At trial, the prosecutors called a college mathematics instructor as an expert witness to quantify the probability of the confluence of a set of traits shared by the interracial couple that committed a Los Angeles robbery and the defendant couple, generating a number approaching a 100 percent probability of guilt.⁵⁹

The California Supreme Court overturned the resulting conviction. In doing so, the court might have taken a narrower route, criticizing the prosecutors for inventing an excuse for emphasizing the one, weak and potentially prejudicial fact they had to go on—that both the perpetrator couple and the defendant couple were interracial. Alternatively, the court might have rested its decision on methodological problems it identified with the state’s amateurish use of statistical evidence: the probabilities used were estimated without data;⁶⁰ the expert failed to show that the frequency of each identifying trait was independent of the frequency of all others, or to acknowledge that his analysis was invalid if the frequencies were not independent (for example, he assigned separate probabilities to the fact that the man had a beard and also a moustache, though the frequency of moustaches is not independent of the frequency of beards);⁶¹ and the expert asked the wrong question (how likely it is that a couple chosen

⁵⁸ *People v. Collins*, 438 P.2d 33 (Cal. 1968).

⁵⁹ A blonde woman with a ponytail stole a woman’s purse, then escaped in a car driven by a black man with a moustache and beard. Neither eyewitness could identify the defendants, Malcolm Collins and his wife, Janet, as the perpetrators. *Id.* at 34. Operating without data, the expert witness estimated frequencies for each trait of the perpetrator couple that matched the defendant couple—e.g., .001 for “interracial couple in car”—and multiplied the frequencies to provide an overall probability that a random couple matching all of the traits would be found in Los Angeles in 1964. The jury found the defendants guilty. *Id.* at 37 & n.10.

⁶⁰ *Id.* at 38.

⁶¹ *Id.*

at random would have the traits of the culprit couple, though the defendant couple were not arrested at random), rather than the correct question (how many such couples there are in the relevant suspect pool—greater Los Angeles).⁶² Instead, the court went on to conclude that, even properly implemented, quantification was a bad idea, because the jury would give undue weight to statistical evidence presented by an expert in numeric form;⁶³ such analysis was beyond the ken of the defense to understand and effectively rebut;⁶⁴ and, most crucially, probabilistic evidence could never answer the question presented, namely, “Of the admittedly few such couples [in Los Angeles], which one, if any, was guilty of committing this robbery . . . in fact[?]”⁶⁵

In the decades since *Collins* was decided, it has been followed, with limited exceptions, in most American jurisdictions.⁶⁶ The decision’s reach and authority were enhanced when the law clerk who assisted in its preparation, Laurence Tribe, joined the Harvard Law faculty and wrote an extensive defense of the decision that added a third argument even more sweeping than the others.⁶⁷ In Tribe’s view, any assault on the admitted myth that trials can achieve certainty, and any quantification of even a high probability of guilt—and, perforce, any acknowledgement of even a minuscule probability of a convicted defendant’s innocence—violates our system’s deepest commitments to “fairness,” trial by jury, the adversarial system and much else that our criminal justice system holds dear.⁶⁸ Tribe’s 1971 article extended this critique beyond the multiplication rule used in *Collins* to a proposal to avoid the foibles of that rule by using more sophisticated, Bayesian analysis.⁶⁹ Since then many other scholars have debated,

⁶² Id. at 42-43; see supra note 46 and accompanying text.

⁶³ *People v. Collins*, 438 P.2d at 40 (“Confronted with an equation which purports to yield a numerical index of probable guilt, few juries could resist the temptation to accord disproportionate weight to that index.”).

⁶⁴ Id. at 41.

⁶⁵ Id. at 40 (“[N]o mathematical equation can prove beyond a reasonable doubt (1) that the guilty couple in fact possessed the characteristics described by the People’s witnesses, or even (2) that only [o]ne couple possessing those distinctive characteristics could be found in the entire Los Angeles area.”).

⁶⁶ See, e.g., cases cited infra note 384.

⁶⁷ Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 Harv. L. Rev. 1329 (1970).

⁶⁸ Id. at 1358-78. Cf. id. at 1393 n.2 (“I am, of course, aware that *all* factual evidence is ultimately ‘statistical,’ and all legal proof ultimately ‘probabilistic,’ in the epistemological sense that no conclusion can ever be drawn from empirical data without some step of inductive inference . . .”).

⁶⁹ See id. at 1329-78 (criticizing Bayesian proposal in Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 Harv. L. Rev. 489 (1970)).

and mainly endorsed and extended, Tribe's objections to the use of systematic aggregative analysis to establish identity in criminal process.⁷⁰

B. *Objections to Aggregative Analysis of Matches and Non-Matches*

Tribe's third critique can be subdivided into three parts: First, the use of aggregative methodology offends fundamental principles, such as the presumption of innocence (allegedly compromised by consideration of the opening odds of guilt⁷¹) or the principle that we punish people based only on proof of what they did, not proof that they are members of a class (allegedly compromised by convicting a defendant after acknowledging, say, a 3 percent chance that she is innocent⁷²). Second, even if aggregative analysis has a place in the criminal justice system, it is so complicated and difficult and so likely to be applied incorrectly that we should categorically exclude it.⁷³ Embedded in this critique is the assumption that lay decision makers, the hallmark of our democratic and decentralized system of criminal justice, do not have the ability and cannot be trained to use the information reliably or to keep from being unduly swayed by large numbers.⁷⁴ Finally, we have no hard data on the frequency of most of the possibilities with which criminal jurors must contend (for example, how often men in a particular neighborhood wear brown sweatshirts), so the values we plug into equations will be sheer conjecture or at least not "quantified 'exactly.'"⁷⁵

⁷⁰ For a sampling of the voluminous literature, see *Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism* (Peter Tillers & Eric D. Green eds. 1988); Lea Brilmayer & Lewis Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. Chi. L. Rev. 116 (1978); L. Jonathan Cohen, *Subjective Probability and the Paradox of the Gatecrasher*, 1981 Ariz. St. L.J. 627 (1981); David Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 Ariz. St. L.J. 101 (1979); Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 Harv. L. Rev. 1357 (1985); Daniel Shaviro, *Statistical Probability Evidence and the Appearance of Justice*, 103 Harv. L. Rev. 530, 547-50 (1989); articles cited *infra* note 384. Symposia and exchanges include *Decision and Inference in Litigation*, 13 Cardozo L. Rev. 253 (1991); *Naturalized Epistemology and The Critique of Evidence Theory*, 87 Va. L. Rev. 1491 (2001); *On the Significance of Batting Averages and Strikeout Totals: A Clarification of the Naked Statistical Evidence Debate, the Meaning of Evidence, and the Requirement of Proof Beyond a Reasonable Doubt, and Responses*, 65 Tul. L. Rev. 1093 (1991); *Probability and Inference in the Law of Evidence*, 66 B.U. L. REV. 377 (1986).

⁷¹ See, e.g., Tribe, *supra* note 67, at 1355-67.

⁷² See, e.g., Ronald L. Allen et al., *Evidence: Text, Problems, and Cases* 164-65 (4th ed. 2006); sources cited *infra* notes 109-110 and accompanying text.

⁷³ See Tribe, *supra* note 67, at 1335.

⁷⁴ See *id.* at 1360.

⁷⁵ *People v. Louie*, 205 Cal. Rptr. 247, 262 (Cal. App. 1984) (citation omitted). See, e.g., L. Jonathan Cohen, *Can Human Irrationality Be Experimentally Demonstrated?*, 4 *Behav. & Brain Sci.* 317, 329 (1981); Tribe *supra* note 67, at 1361.

In the next part we show that aggregative analysis has an established place in our criminal justice system; such analysis can be used correctly; and data limitations are exaggerated. One surprising ally in our argument is the adversarial system itself. When nearly the same arguments were initially made against aggregating small probabilities into big numbers in the process of admitting inculpatory DNA evidence,⁷⁶ the adversarial system quickly elicited workable solutions. We suggest that a similar process—along with modern data-mining techniques and new ways of helping lay audiences understand statistical insights—can enable our criminal justice institutions to develop and refine appropriate methods for aggregating the small probabilities associated with non-exclusionary non-matches into valuable proof of innocence.

IV. STANDARD IDENTITY EVIDENCE AS AGGREGATIONS OF MATCHES AND NON-MATCHES

A. *The Compatibility of Aggregative Analysis and Criminal Justice*

Aggregative analysis of matches and non-matches of small bits of non-unique, non-dispositive evidence is not anathema to our criminal justice system but a core feature of accepted modes of proving identity in criminal cases. Indeed, the new “gold standard” of identity evidence—DNA matching—depends on exactly this kind of analysis.

1. *Confessions, Eyewitness Identifications and Fingerprints as Aggregative Evidence.* — Under the usual understanding, police, prosecutors and jurors seeking to identify the perpetrator of a crime work to match a *unique* trait of the culprit to the accused. Fingerprints embossed in blood on a knife embedded in the victim’s heart and ballistics linking the physical properties of the barrel of a gun owned by a suspect to striations on the bullet found in a gunshot wound are familiar examples.⁷⁷ Older examples are eyewitness identifications matching the witness’s memory of the culprit to a suspect, and confessions or

⁷⁶ For early criticism of the use of DNA in court, see *People v. Castro*, 545 N.Y.S.2d 985, 996-99 (Bronx S. Ct. 1989); Eric S. Lander, DNA Printing on Trial, 339 *Nature* 501 (1989); Stephen M. Patton, DNA Fingerprinting: The *Castro* Case, 3 *Harv. J.L. & Tech.* 223 (1990); Peter J. Neufeld & Neville Colman, Science Takes the Stand, 262 *Scientific American* 46 (1990); Janet C. Hoeffel, Note, The Dark Side of DNA Profiling: Unreliable Scientific Evidence Meets the Criminal Defendant, 42 *Stan. L. Rev.* 465 (1990).

⁷⁷ See, e.g., Kathryn E. Carso, Amending the Illinois Postconviction Statute to Include Ballistics Testing, 56 *DePaul L. Rev.* 695, 700, n.43 (2007); Michael J. Saks, Merlin and Solomon: Lessons from the Law’s Formative Encounters with Forensic Identification Science, 49 *Hast. L.J.* 1069, 1101 (1998).

informant testimony matching known facts about the crime to a description of it by someone who claims to have committed it.

In fact, the power of all of this evidence is not due to a single match of a *unique trait* of the criminal. It is a result of the *confluence of many matches of traits that are not unique* to the defendant, no single one of which is dispositive or (often) very interesting. A fingerprint is powerful because the collection of tiny lines and intersections found in a latent partial fingerprint at a crime scene match those taken from a suspect.⁷⁸ Lines and intersections on a bullet embedded in the victim's heart are powerful if they match imperfections on the inside of the barrel of a gun seized from the defendant.⁷⁹ The same goes for a match between multiple features of a burglary committed by a masked intruder and of a known burglar's classically admissible *modus operandi*,⁸⁰ and of intersections between attributes of a series of similar crimes or possibly criminal events and those in the life of a single suspect that are admissible when they are too numerous and unusual to be coincidental.⁸¹

This insight is not new. Centuries ago, English philosopher William Paley noted that a “concurrence of well-authenticated circumstances composes a stronger ground of assurance, than positive testimony, unconfirmed by circumstances, usually affords. Circumstances cannot lie.”⁸² But, as in Paley's case, this claim typically is made to refute the idea that “circumstantial” evidence is inferior to “direct”

⁷⁸ See, e.g., Laura Spinney, *The Fine Print*, 464 *Nature* 344, 346 (2010) (reporting view of forensic experts that the only way to avoid false-positive fingerprint matches is for “fingerprint analysis [to] be interpreted in probabilistic terms,” using data on “how fingerprint patterns vary across populations and how often various components or combinations of components crop up,” for example, that “a particular configuration of bifurcations, ridge endings [is] found in 40% of a given population”); *infra* note 258 and accompanying text.

⁷⁹ See Paul C. Giannelli, *Forensic Science: Why No Research?* 38 *Fordham Urb. L.J.* 503, 513-18 (2010) (arguing that *Daubert* standard for admitting scientific evidence will eventually require use of more rigorous scientific and statistical principles in fingerprint, handwriting, ballistics, document and other forensic analysis).

⁸⁰ See, e.g., *United States v. Mack*, 258 F.3d 548, 554 (6th Cir. 2001) (“Standard conduct, although not particularly unusual by itself, may, in combination, present an unusual and distinctive pattern constituting a ‘signature.’”); *People v. Haston*, 444 P.2d 91, 100 (Cal. 1968) (“[An] inference of identity arises when the marks common to the charged and uncharged offenses, considered singly or in combination, logically operate to set the charged and uncharged offenses apart from other crimes of the same general variety.”).

⁸¹ See Richard O. Lempert et al., *A Modern Approach to Evidence* 351-54 (4th ed. 2011) (discussing admissibility of “common scheme” evidence). James Baldwin's criticism of the use of “common scheme” evidence to convict Wayne Williams as a serial killer of young children in Atlanta gives this article its title. See Baldwin, *supra* note 1.

⁸² William Paley, *The Principles of Moral and Political Philosophy* 551 (1785).

evidence.⁸³ Our point is stronger—that the power of most or all *direct* evidence is due not to its uniqueness but a concurrence of small bits of individually inconclusive evidence.⁸⁴ Illustrating confusion about the “uniqueness,” “directness” and presumptive strength of certain kinds of evidence is the assumption in popular culture that DNA, fingerprint and other forensic “matches,” which conceptually are on the “circumstantial” side of the divide, instead qualify as (presumptively stronger) “direct” or “unique” evidence.⁸⁵

This point doesn’t end at *modus operandi*, “common scheme” and forensic evidence. It extends, as well, to eyewitness identifications, the quintessential direct or unique evidence. Identifications increase in strength as the witness matches more and more of the suspect’s traits to remembered traits of the culprit. That we base our faith in identifications on the ability of the eye to observe many traits at once and of the brain to process them quickly into a single “aha!” conclusion that the third man from the right is the attacker should not obscure the fact that the eye is seeing and brain is assessing the combined effect of many matching features—most of them uninteresting in themselves—and not a single, unique trait.⁸⁶ Similarly, confessions increase in weight with each new match between actions and instrumentalities known to have been associated with a crime and a confessing suspect’s uncontaminated narrative of what

⁸³ See *infra* notes 304-334 and accompanying text (discussing judicial definitions of “direct” and “circumstantial” evidence that suggest a preference for the former).

⁸⁴ See Simon A. Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 *Law, Probability & Risk* 233, 234, 237 (2009) (criticizing forensic experts who claim that forensic markers are unique to individuals, though they actually can only locate a suspect in a larger or smaller class of possible culprits).

⁸⁵ See Simon A. Cole & Rachel Dioso-Villa, *Investigating the ‘CSI Effect’ Effect: Media and Litigation Crisis in Criminal Law*, 61 *Stan. L. Rev.* 1335, 1340-41, 1364 (2009) (noting popular assumption that forensic evidence is the strongest kind of evidence); Peter Tiersma & Mathew Curtis, *Testing the Comprehensibility of Jury Instructions: California’s Old and New Instructions on Circumstantial Evidence*, 1 *J. Ct. Innov.* 231, 257 (2008) (describing studies showing that mock jurors instructed on difference between direct and circumstantial evidence often mischaracterize highly probative circumstantial evidence as “direct”); *infra* notes 308-311 and accompanying text.

⁸⁶ See, e.g., *Galvin v. Eli Lilly & Co.*, 488 F.3d 1026, 1042 (D.C. Cir. 2007) (“[A]ll evidence, even eyewitness testimony, requires drawing inferences; the eyewitness is drawing an inference from his raw perceptions.” (citation omitted)); *Sylvester v. SOS Children’s Vills Ill., Inc.*, 453 F.3d 900, 903 (7th Cir. 2006) (“All evidence is probabilistic, and therefore uncertain; eyewitness testimony and other forms of ‘direct’ evidence have no categorical epistemological claim to precedence over circumstantial or even explicitly statistical evidence.”); *DePass v. United States*, 721 F.2d 203, 207 (7th Cir. 1983) (Posner, J., dissenting) (arguing that “almost all legal evidence, is probabilistic” . . . and “probabilities that are derived from statistical studies are no less reliable in general than” those “derived from direct observation, from intuition, or from case studies of a single person or event”).

happened.⁸⁷ In these examples, as in the more obviously “circumstantial” ones described above, the inference of guilt moves beyond a reasonable doubt because the probability that the many matches are a result of the defendant’s guilt is so high, and the probability that the matches are coincidental is so low.

What our intuitions tell us is powerful evidence because it reveals a match between a unique trait of the culprit and a suspect almost never is powerful because of the uniqueness of the evidential trait. It is powerful because of the unusual *aggregation* of matches of multiple *non-unique*, often quite *common* traits. The power of the evidence derives from how each additional match (and, we will argue, non-match) of non-unique traits increases, indeed multiplies, the probability of guilt (or innocence). We would argue, further, that in order to understand the extent of the multiplier effect, intuitions again fail criminal justice actors, requiring disciplined measures to expose the aggregate power of the multiple matches or non-matches.

Calls for such measures have increased since DNA exonerations revealed the frequency of false confessions and eyewitness identifications.⁸⁸ A central goal of these improvements is to force into the open as many individual details as possible, to allow the most accurate assessment of the number and quality of matches between the content of confessions and eyewitness testimony and what is known about the crime and criminal. For example, proposals to videotape confessions provide a way to catalogue details to which the defendant claims to be confessing to see whether they match the known features of the crime and of verifying that investigators did not feed details to the suspect.⁸⁹ Similarly, proposals to document pre-lineup eyewitness descriptions with sketch artists and cross-reference them to descriptions other witnesses independently gave help to specify and clarify the many separate features that eyewitness

⁸⁷ There is probative value in a confession without details (“I did it”) or a line-up identification by a victim who could not give a description. But as exonerations and empirical research show, a suspect’s willingness to confess or resemblance to a victim’s memory of a criminal—especially when proffered without supporting details—is not a unique trait of guilty people, much less of any single guilty perpetrator. See, e.g., Donald Connery, *Convicting the Innocent: The Story of a Murder, a False Confession, and the Struggle to Free a “Wrong Man”* 148 (1996) (documenting how a suspect’s youth, inexperience with the justice system, mental illness and suggestibility disposed him to confess falsely); Steven A. Drizin & Richard A. Leo, *The Problem of False Confessions in the Post-DNA World*, 82 N.C.L. Rev. 891 (2004) (noting that innocent suspects sometimes confess to stop exhausting interrogations, to please authorities or because of youth or mental disability); Sarah Anne Mourer, *Reforming Eyewitness Identification Procedures Under the Fourth Amendment*, 3 Duke J. Const. L. & Pub. Pol’y 49, 56 (2008) (discussing disposition of eyewitnesses to make inaccurate identifications to provide closure).

⁸⁸ See, e.g., Garrett, *supra* note 18, at 42 (cataloguing confessions and identifications later shown to be false).

⁸⁹ For articles advocating the videotaping of interrogations, see *supra* note 23; *infra* note 446.

identifications typically conflate, exposing the number and weight of each match and non-match.⁹⁰ Although measures of this sort are far from universally accepted, they are becoming more common, and no one suggests they are inconsistent with basic criminal justice norms.⁹¹

2. *The Courts' Enthusiastic Embrace of Aggregative Analysis: Inculpatory DNA.* — Even more clearly establishing the centrality to modern criminal justice of the systematic, mathematical, aggregation of the probative value of multiple non-unique traits is inculpatory DNA, the *strongest* known form of identity evidence. Using none other than the multiplication rule pilloried in *Collins*,⁹² inculpatory DNA derives its power entirely through the aggregation of individually unimpressive probabilities of guilt associated with multiple shared traits. Although absorbed into the public consciousness as “DNA fingerprinting,”⁹³ appropriating the common misperception of fingerprint evidence as unique to a single person,⁹⁴ inculpatory DNA evidence does not depend on any unique feature of individuals.⁹⁵ Instead, it relies on a confluence of matches of individual DNA sequences, each of which recurs in the relevant population with a known and considerable frequency.

Modern DNA profiling uses a Polymerase Chain Reaction (PCR), followed by gel or capillary electrophoresis, to identify DNA sequences known as short tandem repeats (STR).⁹⁶ STRs are highly variable regions of the human genome that are made up of short repeating sequences of the nucleotides that comprise DNA. For example, one STR is be made up of the short nucleotide sequence CATG that repeats over and over again (e.g., CATGCATGCATGCATG). A given STR is found at a discrete physical location (“locus”) in the human genome. Although that locus exists on everyone’s individual genome, the number of times the nucleotide sequence repeats at that STR locus (in our example, the

⁹⁰ See Neil Vidmar, Rethinking Reliance on Eyewitness Confidence, 94 *Judicature* 16, 17 (2010). For parallel proposals in the fingerprint context, see *supra* note 78.

⁹¹ See *supra* note 23 (noting increasing number of jurisdictions that videotape confessions).

⁹² See *supra* notes 58-65 and accompanying text.

⁹³ See, e.g., Lander, *supra* note 76, at 505; William C. Thompson and Simon Ford, DNA Typing: Acceptance and Weight of the New Genetic Identification Tests, 75 *Va. L. Rev.* 45, 49 (1989).

⁹⁴ See *supra* note 78.

⁹⁵ Although, with the exception of monozygotic twins, each individual has a unique genome, forensic DNA profiling cannot yet examine enough of an individual’s entire genome to establish uniqueness. See Harlan Levy, *And the Blood Cried Out* 24 (1996) (explaining that testimony in connection with DNA evidence is “highly misleading if it leads us to believe that the power of DNA analysis lies in its ability to identify . . . characteristics . . . unique to each individual. Science does not yet have that power.”).

⁹⁶ David H. Kaye, *The Double Helix and the Law of Evidence* 187 (2010).

number of CATGs in a row) varies among individuals. One person may have five repeats; another may have six, and so on. Each different numbers of repeats is called an “allele.”⁹⁷ Variation in the alleles found at each STR locus can help discriminate among individuals.⁹⁸

Individually, however, each allele is fairly common. Often as many as twenty percent of the population share a given allele.⁹⁹ As a result, the presence of a single match between an allele in DNA left at a crime scene and in the DNA of a suspect is unremarkable and no more discriminating than, for example, the fact that both the mid-twenties criminal and mid-twenties suspect have male pattern baldness or are left-handed.¹⁰⁰ When alleles from *multiple* locations match, however, the power of DNA evidence emerges. More exactly, the probability that a suspect arrested at “random”—i.e., for reasons other than his genetic make-up—would have the matching number of repeats at each of multiple STR loci is the product of the frequencies of each, individually rather common, allele.¹⁰¹ The more STR regions that are tested in an individual sample, the more discriminating the test becomes. If matching STRs appear at thirteen different loci (the number of loci typically examined in forensic DNA testing by the FBI¹⁰²), and if the incidence of each STR is 20 percent (making each trait more common than left-handedness and male-factor baldness in twenty-five-year-olds), the probability of selecting an individual at random with the same collection of STRs is less than one in many trillions, as the jury will be informed.

Going beyond the multiplication rule, several US courts in civil and criminal paternity cases have endorsed Bayes’ Theorem as the most acceptable way to update the “prior odds” estimate of paternity

⁹⁷ Michael J. Saks et al., *Annotated Reference Manual on Scientific Evidence* 492, 494 (2d ed. 2005) (noting that each individual has two alleles at a given STR locus, one from each chromosome—e.g., one that codes for green eyes and one for blue eyes).

⁹⁸ *Id.* at 188.

⁹⁹ See Hoeffel, *supra* note 76, at 538.

¹⁰⁰ “A typical allele might be found in 10 percent of the population, making it not all that unlikely that two random people will carry the same allele. But if one looks at alleles at [multiple] sites, [a match] becomes increasingly unlikely . . .” Neufeld & Colman, *supra* note 76, at 50. See also D. C. Gan & R. D. Sinclair, *Prevalence of Male and Female Pattern Hair Loss in 3 Maryborough J. Investig. Dermatol. Symp. Proc.* 184-89 (Dec. 2005) (reporting that approximately 17 percent of men in their twenties exhibit symptoms of male pattern baldness); M. Raymond et al., *Frequency-Dependent Maintenance of Lefthandedness in Humans*, 22 *Proceedings Biol. Sci.* 1627 (1996) (estimating frequency of lefthandedness as 10-13 percent).

¹⁰¹ In DNA analysis, the “Basic Product Rule”—the multiplication rule applied to DNA analysis—requires an estimate of the frequency of the alleles at each locus (the “single locus genotype”), then multiplies the figures to estimate the frequency of the combination of single-locus genotypes (“multi-locus genotype”). See Kaye, *supra* note 96, at 89-90.

¹⁰² *Id.* at 189.

based on non-forensic evidence with likelihood-ratio probabilities associated with evidence of the defendant's DNA profile—i.e., the probability of the child's genetic profile if the defendant is his father divided by the probability of the profile if someone else is his father.¹⁰³ The cases have triggered a healthy debate about how best to estimate the prior probability of paternity based on non-scientific evidence—for example one over the number of men with whom the mother may have had sexual relations around the time the child was conceived.¹⁰⁴

We thus can add inculpatory DNA to our list of allegedly “unique” evidence that inculpates not by way of a single match of a unique trait exhibited by the perpetrator and suspect (e.g. an individual's genome) but by way of a confluence of many matches (e.g., traits found at various loci on the suspect's chromosomes). No single loci-match is dispositive or even strongly indicative of the defendant's guilt, but taken together, multiple loci-matches produce a high probability—though still not a certainty—that the defendant and the donor of the material found at the scene are the same person. In *principle*, there is no reason why equally powerful results cannot emerge from collections of other, individually unimpressive, matching traits of known frequency, such as height, handedness, hair loss, eye color, ear lobe configuration, etc. By analogy to the DNA paternity cases, Bayes Theorem, as well as the multiplication rule, could also be available for these purposes.

There is, to be sure, a difference in *practice* between how jurors (as well as police and prosecutors) experience a match of DNA alleles and how they experience a confluence of other matching traits. In the case of DNA, an expert bio-statistician will represent the aggregate strength of the string of matches as a number—the product of the probabilities associated with each matching allele. An expert, that is, will help the jury or other criminal-justice actor to understand not only that each new item in a series of matching traits increases the probability of guilt, but also the *magnitude* of the “*f*multiplier” effect of each new match. In other cases, no such expert assistance will be provided. Even if the data

¹⁰³ See, e.g., *In re the Paternity of M.J.B.*, 425 N.W.2d 404, 408-09 (Wis. 1988) (describing various statistical techniques).

¹⁰⁴ See 4 Faigman et al., *supra* note 29, at 189-96 (describing use of Bayes' Theorem in paternity cases); David H. Kaye et al., *The New Wigmore: Expert Evidence* 478-91 (2004) (same); *infra* notes 152-156 and accompanying text (same).

needed to calculate the relevant numbers are available—as in the case of a short, green-eyed, left-handed, balding culprit—the *Collins* line of cases will prevent the prosecutor from calling (and probably dissuade her, outside of court, from consulting) a statistician to report a number.

Indeed, unlike in the case of forensic evidence (for example, blood types and DNA profiles), for which the courts have grown increasingly comfortable with testimony quantifying the probability of a random match,¹⁰⁵ courts often exclude evidence of the frequency of other identifying traits—a prerequisite to systematically aggregative analysis of the conjoint power of several such traits. Unless the reference class for a frequency study of a non-forensic trait “exactly” matches the facts of the case at hand, many courts bar evidence of the frequency—for example, how often particular medical diagnoses are made (of interest in insurance fraud cases); the probability of arranging a list of ten companies in the same, non-alphabetical order to prove that a search warrant, the application for which arranged the companies in that order, was the forbidden fruit of a prior, unlawful search; or the frequency of mistaken eyewitness identifications.¹⁰⁶ As Professor Koehler has pointed out, such “reference class requirements . . . are so extreme that they would eliminate the use of statistical evidence under nearly all conditions,” including inculpatory DNA itself.¹⁰⁷

We doubt that the distinctions courts currently draw among identifying traits for purposes of quantifying and systematically aggregating random-match probabilities will hold up over time. Even if some such distinctions make sense, they are hardly stable and fundamental enough to sustain Tribe’s and others’ claim that quantification and aggregation threaten the foundations of our criminal justice

¹⁰⁵ See, e.g., *People v. Mountain*, 486 N.E.2d 802, 804 (N.Y. 1985) (reversing prior precedent and allowing introduction of forensic blood-type matches and associated random-match frequencies); Jonathan J. Koehler, *When Do Courts Think Base Rate Statistics Are Relevant?*, 42 *Jurimetrics J.* 373, 388-89 (2002) (“Almost all courts find [forensic science base rates] to be relevant and admissible.”); Dale A. Nance, *Naturalized Epistemology and the Critique of Evidence Theory*, 87 *Va. L. Rev.* 1551, 1610 (2001) (“Most criminal courts . . . allow the presentation of the random match statistic [i.e., the frequency of a forensic trait in the general population]”); 1 Faigman et al., *supra* note 29, § 15-5.3 (citing cases).

¹⁰⁶ See Koehler, *When Do*, *supra* note 105, at 384-85, 390-93 (citing cases). But see *id.* at 380-90 (citing decisions allowing statistical evidence of (1) the small probability of Sudden Infant Death Syndrome or (2) a chance collection of cardiac arrests to establish criminal or civil liability for multiple deaths in rapid succession in the same family or hospital, and of (3) the small probability of a particular pattern of accurate and inaccurate answers to prove cheating on a test).

¹⁰⁷ *Id.* at 392. The reference group the FBI and other agencies use to reveal the frequency of particular STRs is not a random sample of any population, much less the population of potential suspects in any given crime. See sources cited *infra* note 164. DNA evidence thus cannot satisfy the “reference class” requirements that bar aggregative analysis of other evidence.

system.¹⁰⁸ At the heart of the critique is the prediction that parties and the public will never accept verdicts that are or appear to be based on the defendant's membership in a class—even a very small class—of people who could be guilty, as opposed to a determination that the defendant is “uniquely” guilty.¹⁰⁹ The verdict, it is claimed, must convey the message that the defendant *is*, not that she “probably” or “almost certainly” is, guilty.¹¹⁰

Our courts, however, have long acknowledged that criminal verdicts are “merely” probabilistic. For years, judges have instructed jurors that proof beyond a reasonable doubt does not mean proof beyond *any* doubt or to an “absolute . . . certainty” and that they may find the defendant guilty though they are “fully aware” that their verdict is based on “probabilities” and “may be mistaken.”¹¹¹ Inculpatory DNA cements the point. DNA achieves its status as the “gold standard” of proof of identity by mathematically aggregating probabilities associated with a series of individually inconclusive matching traits to an overall probability less than one that the defendant left the genetic material at the crime scene. Recently, the Supreme Court even suggested that it would be error to mislead jurors about two statistical realities of DNA evidence: (1) it can reveal only a probability less than one that the defendant or the victim is the source of biological material found at the crime scene or on the defendant.¹¹² (2) The probability that the defendant is *guilty* is even lower than that, given a probability greater than zero that investigators erred in collecting or analyzing the evidence and that there is an innocent reason why the defendant's or victim's genetic material was found in a seemingly incriminating location.¹¹³ Despite these realities, courts on both

¹⁰⁸ See *supra* notes 68-75 and accompanying text.

¹⁰⁹ See Nesson, *supra* note 70, at 1378 (“[A] probable verdict may not be acceptable [to the public], and an acceptable verdict may not be probable.”); Tribe, *supra* note 67, at 1372, 1375 (claiming that probabilistic evidence “dehumanizes” justice and weakens public and party support for the legal system).

¹¹⁰ See Nesson, *supra* note 70, at 1390 (arguing that the goal of trials is to project behavioral norms to the public by linking authoritative narratives about what happened to legal consequences, and that probabilistic verdicts, no matter how accurate, undermine this goal); Tribe, *supra* note 67, at 1372 (worrying that evidence of probabilities less than one in criminal trials could undermine the policy of acquittal in the face of reasonable doubt).

¹¹¹ See, e.g., *Victor v. Nebraska*, 511 U.S. 1, 31 (1994) (upholding constitutionality of jury instruction using this language).

¹¹² See *McDaniel v. Brown*, 130 S. Ct. 665, 670 (2010) (discussed *infra* notes 147-151 and accompanying text).

¹¹³ See *id.* at 671 (discussed *infra* notes 145-151 and accompanying text); Norman Fenton & Martin Neil, *Avoiding Legal Fallacies in Practice Using Bayesian Networks*, Seventh International Conference on Forensic Inference and Statistics (Aug. 2008), at 7 (“Errors in the DNA typing process can result in a reported match where there is not true match. A true match can be coincidental if more than one member of the population shares the DNA features recorded in the sample; . . . even if the defendant was the source he/she may not be the perpetrator since there may be an innocent reason for their presence at the crime

sides of the Atlantic have no problem upholding convictions based on little more than a DNA match (actually, a series of mathematically aggregated matches).¹¹⁴

Clearly, our criminal justice system exhibits no blanket preference for a false certainty over estimated probabilities as a matter of liberal democratic fundamentals, nor should it when life and liberty depend on the accuracy of verdicts. Nor are plain statements in trials about the probabilistic nature of verdicts—and plain demonstrations of the ability to build a powerful case of guilt or innocence by systematically aggregating individually unimpressive probabilities associated with each of a string of matching or non-matching traits—in any way subversive of the our justice system or dissuasive of party participation in and public acceptance of the system.

To be sure, as we discuss in Part V, human beings do seem to exhibit a *cognitive* preference for a false certainty over accurately estimated probabilities.¹¹⁵ That cognitive bias, however, can lead us to accept demonstrably false things as true, and arises not because we prefer falsehood over truth but because we “prefer” the false sense of comfort that our brains trick us into associating with certainty, however derived.¹¹⁶ Assuming that, instead of mimicking our cognitive foibles, the justice system should help overcome them—achieving through legal tools what we cannot accomplish with our bare hands and brains—its goal should be to process *all* the information we have, aggregated probabilities included, without generating more anxiety than we assuage by knowing we got the best possible answer.¹¹⁷ We see nothing in the fundamentals of our justice system that is inconsistent with the goal of improving upon practical psychology.

B. The Adversarial System’s Ability to Domesticate Aggregative Analysis

scene.”); Jonathan J. Koehler, Error and Exaggeration in the Presentation of DNA Evidence at Trial, 34 *Jurimetrics* 21 (1993) (similar).

¹¹⁴ See, e.g., *Spencer v. Commonwealth*, 384 S.E.2d 775, 782 (Va. 1989) (affirming capital verdict premised mainly on a match between Spencer’s genetic profile and that in semen at the crime scene and on testimony that the probability of a random match was 1/135,000,000). Spencer was executed on April 27, 1994. See <http://www.deathpenaltyinfo.org/executions>. See also *supra* notes 43-57 and accompanying text (discussing *R. v. Adams*).

¹¹⁵ See *infra* note 218-233 and accompanying text.

¹¹⁶ See *infra* note 234-236 and accompanying text.

¹¹⁷ See generally Gerd Gigerenzer, *Reckoning with Risk: Learning to Live with Uncertainty* (2002).

The question remains, however, whether systematically aggregated probabilities associated with non-exclusionary non-matches can help reach accurate answers, rather than confusing or misleading decision makers. Here again, we argue that inculpatory DNA provides much of the answer. As we have seen, inculpatory DNA combines strategic data mining to ascertain the frequency of multiple non-unique identifying traces associated with a crime and statistical methodology for aggregating those individually rather high probabilities into a small chance of an accidental match between the traces and a defendant.¹¹⁸ In view of how often exonerations highlight potentially identifying traces found during police investigations that did *not* match the wrongly convicted defendant but turned out to match the actual perpetrator,¹¹⁹ we propose a similar combination of data mining and statistical analysis (in this case Bayes' Theorem) to aggregate the individually small probabilities of innocence associated with each of a series of non-exclusionary non-matches into large enough probabilities to raise a reasonable doubt about guilt.

Merely imagining this proposition calls to mind serious technical questions akin to those that provided a sufficient and uncontroversial basis for the *Collins* decision:¹²⁰ Even with expert assistance, how can lay decision makers reliably recognize non-matches, assign independent probabilities to them and aggregate those probabilities with prior odds of guilt that are themselves of uncertain provenance and independence? In fact, DNA's use to establish criminal identity and paternity raises the same concerns, which over time the adversarial system has worked hand in glove with the relevant technologists (biologists and statisticians) to allay. We suggest that the same give and take between proposed methods, adversaries' objections and responsive improvement of the state of the art can work in this new context as well.¹²¹

¹¹⁸ See supra notes 92-102 and accompanying text.

¹¹⁹ See supra notes 24-26 and accompanying text.

¹²⁰ See supra notes 60-62 and accompanying text.

¹²¹ See Kaye, supra note 96, at 58-160 (describing early 1990s disputes over admissibility of DNA evidence); Jennifer L. Mnookin, *The Image of Truth: Photographic Evidence and the Power of Analogy*, 10 *Yale J.L. & Human.* 1, 70 (1998) (discussing analogous process through which courts initially resisted, then accepted photographic evidence).

Although celebrated today, inculpatory DNA analysis initially was crude and controversial. In *People v. Castro*,¹²² an early case in which the technique was mooted in court, Dr. Richard Roberts, a molecular biologist in the lab of DNA pioneer and Nobel Prize Winner James Watson,¹²³ validated other prosecution testimony placing “the odds of a random match between a bloodstain [on Castro’s watch band] and the [genetic profile of the badly beaten murder victim] at one in 100 million.”¹²⁴ Testifying for the defense, Eric Lander, a Harvard and MIT mathematician, scientist and MacArthur award recipient,¹²⁵ “examined the same data and arrived at odds of one in 24.”¹²⁶ After appearing as opposing experts, Roberts and Lander met privately while the case was in recess, decided they both were wrong in part and jointly authored a statement that was subsequently introduced in court. The statement called for further testing and analysis because “the DNA in this case are not scientifically reliable enough to support the assertion that the samples . . . do or do not match.”¹²⁷ In a carefully reasoned decision in August 1989, the trial judge concluded that DNA matching is a potentially reliable form of identity evidence but was inadmissible in the case due to methodological improprieties in data mining and statistical analysis that were described in detail.¹²⁸ Later that year, Dr. Lander published an article in the prestigious scientific journal *Nature* cataloging defects in the proposed use of DNA in the *Castro* case and proposing improvements in data mining techniques, statistical analysis and courtroom procedures to avoid the

¹²² 545 N.Y.S.2d 985 (Bronx Cnty. S. Ct. 1989).

¹²³ See Levy, supra note 95, at 42 (describing *Castro* trial).

¹²⁴ Neufeld & Colman, supra note 76, at 47-48.

¹²⁵ See Levy, supra note 95, at 43. See also Gina Kolatta, Power in Numbers, N.Y. Times, Jan. 2, 2012, http://www.nytimes.com/2012/01/03/science/broad-institute-director-finds-power-in-numbers.html?pagewanted=1&_r=1&ref=science?src=dayp (profiling Eric Lander).

¹²⁶ Neufeld & Colman, supra note 76, at 48.

¹²⁷ Lander, supra note 76, at 504 (citation omitted). See Levy, supra note 95, at 47.

¹²⁸ *People v. Castro*, 545 N.Y.S.2d at 996 (criticizing prosecution’s DNA evidence for failing to use proper probes when analyzing degraded biological samples; include male and female controls when testing for sex of sample on defendant’s watch; and test for non-human DNA in samples, rather assuming all strands were of human origin).

problems in the future.¹²⁹ A 1990 *Scientific American* article by one of Castro's lawyers, Peter Neufeld, presented an overlapping set of critiques and methodological and procedural solutions.¹³⁰

The defects identified in early cases and commentaries fell into four categories: (1) Chemical, autoradiograph and other data-mining techniques were not discriminating enough to identify exactly what alleles—or other genetic traces or contaminants—were present in samples from the crime and defendant.¹³¹ (2) Absent accepted standards, determinations of whether autoradiographed reproductions of the same allele in the two samples were sufficiently clear and similar in length to establish a match were unreliably subjective.¹³² (3) Technicians used invalid statistical methods to determine the chance of a random match, including the failure to sample enough human subjects to generate reliable frequencies of particular alleles in different populations;¹³³ the segmentation of reference samples only by race (e.g., African-American) or ethnicity (e.g., Hispanic), and not by potentially isolated subgroups within each race or ethnicity, creating a risk of non-independence among different alleles found in both samples (e.g., Afro-Cubans might have a higher incidence of clusters of particular alleles than blacks or Hispanics as a whole);¹³⁴ improper assumptions about the source population for the trace found at the scene (e.g., the assumption that a crime occurring in a predominantly Hispanic part of the Bronx must have been committed by a Hispanic defendant); and the failure of private laboratories to reveal “proprietary” assumptions about the frequency of alleles in different populations, preventing cross-laboratory

¹²⁹ See Lander, *supra* note 76, at 501-04 (criticizing use of visual, rather than objective, thresholds for declaring matches between crime-scene and suspect samples; poor documentation of size of control-groups; conclusions drawn from a single degraded band of DNA; avoidable risk of contamination from foreign DNA; reliance on assumptions about population genetics that, e.g., ignore heterogeneity in Hispanic populations).

¹³⁰ See, e.g., Neufeld & Colman, *supra* note 76, at 50-53 (advocating improved methods to determine whether allele from one sample is the same as that from another, especially with small samples that prevent retesting; avoid misjudgments due to contamination of crime-scene specimens, bacterial degradation, improperly prepared gels and over-concentrated samples without available control groups; standardize sampling frequencies and population estimates across laboratories; and determine whether differences in allele frequency among ethnic subgroups invalidate statisticians' assumption of random mating across racial groups).

¹³¹ See, e.g., *People v. Castro*, 545 N.Y.S.2d at 991; Kaye, *supra* note 96, at 51; Lander, *supra* note 76, at 502-03; Neufeld & Colman, *supra* note 76, at 51.

¹³² See, e.g., *People v. Castro*, 545 N.Y.S.2d at 994; Kaye, *supra* note 96, at 95-96; Lander, *supra* note 76, at 502-03; Neufeld & Colman, *supra* note 76, at 51; Hoeffel, *supra* note 76, at 479.

¹³³ See Patton, *supra* note 76, at 236 n.41.

¹³⁴ See, e.g., *People v. Castro*, 545 N.Y.S.2d at 992; Kaye, *supra* note 96, at 124; Lander, *supra* note 76, at 504; Patton, *supra* note 76, at 236; Hoeffel, *supra* note 76, at 489-90.

comparison and peer review. (4) Laboratories failed to document instances and patterns of shoddy or improper techniques in collecting, handling and analyzing samples.¹³⁵

In all four cases, public and private technicians charged with collecting and analyzing samples and attaching probabilities to them, and prosecutors and judges responsible for their handling in court, quickly set about solving or debunking the problems. Their efforts were aided by panels convened by the National Research Council, the National Academy of Science and other independent bodies.¹³⁶ Under constant adversarial scrutiny from defense lawyers in court, following Peter Neufeld and Barry Scheck's lead in *Castro*, and to allay doubts expressed in scientific and legal publications and the press,¹³⁷ law enforcement and the courts developed workable solutions to all of these problems and others recognized later.

Data mining techniques have drastically improved. The initial method of isolating alleles for measurement and comparison were based on an unproven assumption that particular chemical probes would isolate material at only a single locus on the crime scene and suspect's samples and thus that the radiographs taken of the two samples could be compared to see if they matched. Responding to defense and judicial concerns, analysts developed new chemical probes that demonstrably would bind only with a single locus, assuring for the first time that radiographs generated by each of the two samples permitted apples-to-apples comparisons.¹³⁸ Additionally, more automated and repetitive techniques eliminated much of the subjectivity associated with previous methods of determining whether the gene sequences in the two pictures were sufficiently distinct and similar in length to establish a match and required less genetic material, permitting independent replication in more cases.¹³⁹

¹³⁵ See, e.g., *People v. Castro*, 545 N.Y.S.2d at 997; Kaye, *supra* note 96, at 101; Neufeld & Colman, *supra* note 76, at 53; Hoeffel, *supra* note 76, at 493-95.

¹³⁶ See National Research Council, Committee on DNA Technology in Forensic Science, *DNA Technology in Forensic Science* (1992) (hereinafter NRC I); National Research Council, Committee on DNA Technology in Forensic Science: An Update, *The Evaluation of Forensic DNA Evidence* (1996) (NRC II).

¹³⁷ See Levy, *supra* note 95, at 49 ("To this day, much of the lingering popular perception that DNA evidence is somehow flawed has its origin in the press coverage of the *Castro* case); sources cited *supra* notes 76, 113.

¹³⁸ See, e.g., Kaye, *supra* note 96, at 50.

¹³⁹ See *id.* at 179.

Adversarial and scientific criticism and judicial oversight also prompted solutions to most of the problems in the third, statistical, category. Many more reference samples were collected, and the same standard was used to confirm the presence of an allele in members of reference populations as in comparing crime-scene and suspect samples, making random-match statistics more reliable, transparent and uniform across labs.¹⁴⁰ Lawyers and researchers continue to disagree over the best way to adjust random-match probabilities based on the size of the database of reference samples, but several solutions are currently in use.¹⁴¹

Interdependence worries based on the possibility that particular alleles and clusters of them are more common in racial and ethnic sub-groups than within racial and ethnic populations as a whole took longer to dispel. Initially, analysts adopted a conservative estimation technique called “ceiling frequencies,” which used the highest estimate of the frequency of each allele in any known sub-population as the frequency for all populations, so the actual aggregate probability of a random-match was likely to be much lower in fact than the conservative probability offered in court.¹⁴² Simultaneously, researchers experimented to see if greater specificity in reference samples—e.g., sampling individuals of Afro-Cuban descent, instead of only sampling blacks or Hispanics as a whole—changed the estimates of the frequency of particular alleles and clusters in particular populations. Soon, “specificity concerns were allayed by empirical demonstrations that increases in reference class specificity made little difference.”¹⁴³

Most controversies now are in the final, more mundane category of faulty implementation of accepted data-mining and statistical methods. An example is the so-called “prosecutor’s fallacy,” which occurs when a prosecutor or witness transforms the conditional probability of a *random match* (say,

¹⁴⁰ See *id.* at 138.

¹⁴¹ See *id.* at ____.

¹⁴² See NRC I, *supra* note 136, at 82-85 (concluding that ceiling frequencies are a scientifically acceptable alternative to sampling every conceivable subpopulation); NRC II, *supra* note 136, at 5-8 (identifying database alternatives when a particular subpopulation has no available sample set).

¹⁴³ Koehler, *When Do*, *supra* note 105, at 393-94 (noting that in the 1990s, defense lawyers often challenged DNA random-match probabilities (RMPs) on the ground that DNA base rates derived from reference classes “did not account for substructuring, i.e., variability in the frequency of genetic profiles across ethnic subgroups,” but that “research quickly convinced most scientists that substructure affected RMPs in [only] minor ways”); Eric Lander & Bruce Budowle, *DNA Fingerprinting Dispute Laid to Rest*, 371 *Nature* 735 (1994) (similar).

1/100) into the probability of the *defendant's guilt* (99 percent).¹⁴⁴ This transposition ignores the prior odds. Even if 99 percent of all lawyers carry briefcases and other people rarely do, the probability that a randomly selected briefcase carrier is a lawyer is much less than 99 percent, given the vast proportion of the population made up of (albeit infrequently briefcase-encumbered) non-lawyers. To figure out how likely it is that a given brief-case carrier is a lawyer requires us to know, as well, the “prior likelihood of being a lawyer. Similarly, to draw conclusions about the probability [that] a criminal suspect is guilty based on evidence of a [DNA] ‘match,’ we must consider not just the percentage of people who would match but also the prior likelihood that the defendant in question is guilty.”¹⁴⁵ That requires consideration of the number of possible suspects whose biological profile is unknown and the strength of the other evidence or guilt or innocence. It thus takes “Bayes’ Theorem . . . to calculate the amount one should revise one’s estimate of the probability of a suspect’s guilt after receiving [DNA] evidence accompanied by incident rate statistics.”¹⁴⁶ The Supreme Court’s recent recognition that allowing jurors presented with DNA to operate under the prosecutor’s fallacy when evidence may be fundamentally unfair will no doubt accelerate the search for solutions, including the routine use of Bayes’ Theorem to highlight the role of prior odds.¹⁴⁷

Bayes’ Theorem helps formalize two related problems: Properly calculating the likelihood-ratio denominator—the probability that a match would appear though the defendant is innocent—requires consideration not only of the random-match probability but also (1) the frequency of laboratory false

¹⁴⁴

[I]f a juror is told the probability a member of the general population would share the same DNA is 1 in 10,000 . . . , and he takes that to mean there is only a 1 in 10,000 chance that someone other than the defendant is the source of the DNA found at the crime scene . . . [,] he has succumbed to the prosecutor’s fallacy.

McDaniel v. Brown, 130 S. Ct. 655, 670 (2010). See e.g., William C. Thompson & Edward L. Schumann, Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor’s Fallacy and the Defense Attorney’s Fallacy, 11 L. & Hum. Behav. 167, 169-71 (1987).

¹⁴⁵ Thompson & Schumann, *supra* note 144, at 170 & n.2.

¹⁴⁶ *Id.*

¹⁴⁷ See McDaniel v. Brown, 130 S. Ct. at 665, 670 (acknowledging that prosecutor’s resort to the prosecutor’s fallacy in closing argument could mislead jurors, potentially making the trial “fundamentally unfair,” but denying relief because issue had not been properly raised below); Christopher M. Triggs & John S. Buckleton, Why the Effect of Prior Odds Should Accompany the Likelihood Ratio When Reporting DNA Evidence, 3 L. Prob. & Risk, 73, 76 (2004).

positives¹⁴⁸ and (2) the possibility that the defendant's DNA was at the scene for reasons other than his commission of the crime, for example, that he lived there or was framed.¹⁴⁹ Courts now address the first part of the problem by admitting evidence of error rates for particular laboratories and analysts.¹⁵⁰ The Supreme Court's recent recognition that it is "error" for prosecutors to equate the probability of guilt with the random match probability without taking account of the possibility of innocent or frame-up reasons for the presence of the defendant's DNA will no doubt motivate solutions to the rest of the problem, again including use of Bayesian analysis to highlight the key considerations.¹⁵¹

Nor is it fanciful any longer to contemplate the routine use of Bayes' Theorem in criminal cases. As we note above, a combination of DNA and Bayesian analysis presented by experts has recently revolutionized proof of paternity,¹⁵² including proof of identity in rape cases involving minors or severely disabled nursing home patients who give birth to a child.¹⁵³ DNA analysis can provide solid numbers for the numerator and denominator of the likelihood ratio in the paternity context but cannot provide the prior odds of paternity. The jury must estimate those odds using non-scientific evidence of the number of possible sexual partners of the mother of the child in question and other information pointing to one possible partner or another. To solve the "prior odds" problem in order to make full use of Bayesian analysis, courts have adopted several competing strategies, including allowing experts to report results based on 50-50 prior odds, while permitting the opposing side to offer expert testimony or requiring the trial judge to give instructions inviting the jury to alter the 50-50 assumption;¹⁵⁴ allowing experts for each

¹⁴⁸ See Koehler, *When Do*, supra note 105, at 394 & n.118 (citing laboratory error rates ranging from 1 out of 67 to 1 out of 345 false positives).

¹⁴⁹ See Jonathan J. Koehler, *Why DNA Likelihood Ratios Should Account for Error (Even When a National Research Council Report Says they Should Not)*, 37 *Jurimetrics J.* 425, 432 (1997).

¹⁵⁰ See Koehler, *When Do*, supra note 105, at 394-95 & nn.119, 124 (citing cases); Thompson & Schumann, supra note 144, at 177 (similar).

¹⁵¹ *McDaniel v. Brown*, 130 S. Ct. at 670 ("It is . . . error to equate source probability with probability of guilt, unless there is no explanation other than guilt for a person to be the source of crime-scene DNA."); see articles cited supra notes 145-149 (advocating full use of Bayesian analysis).

¹⁵² See supra notes 103-104 and accompanying text.

¹⁵³ See, e.g., the *Griffith* decision cited infra note 154.

¹⁵⁴ See, e.g., *Davis v. State*, 476 N.E.2d 127, 138 (Ind. Dist. Ct. App. 1985) (affirming convictions premised on proof of parental relationship to victim which was established using Bayes' Theorem and 50-50 "prior odds"); *Kammer v. Young*, 535 A.2d 936, 940 (Md. Ct. Spec. App. 1988) (permitting use of 50-50 prior odds as long as defendant could offer evidence and argument attacking that assumption); *Commonwealth v. Beausoleil*, 490 N.E.2d 788, 797 & n.19 (Mass. 1986) (similar); *Griffith v. State*,

side to propose prior odds based on their understanding of the evidence and letting the jury choose between the competing estimates or make its own;¹⁵⁵ or providing the trier of fact with a chart indicating how any prior odds the jury might estimate based on non-DNA evidence—from 1:99 (1%) to 99:1 (99%)—would interact with the scientifically generated likelihood ratio to produce specified subsequent odds of paternity.¹⁵⁶

It thus has taken only twenty years of trial, quickly recognized error and responsive refinement for the adversarial system to motivate scientists, lawyers and judges to devise workable solutions to the data-mining and statistical problems that initially threatened the viability of DNA evidence.¹⁵⁷ Statistical hurdles overcome include several that *Collins* and Professor Tribe treated as nearly insurmountable obstacles to systematically aggregating probabilities associated with identifying traits.¹⁵⁸ Solutions include the conservative estimation of frequencies to mitigate the interdependence problem, the “chart” strategy for helping jurors integrate subjective prior odds with “harder” data-mined probabilities and comparative frequency testing of different reference groups to determine the point where further sub-grouping does not much improve reliability.¹⁵⁹ Because these solutions apply to the systematic aggregation of probabilities associated with all identifying traits, there is no reason to think they would not be serviceable, and further improve in response to adversarial pressure, outside the DNA context.

976 S.W.2d 241, 245-46 (Tex. App. 1993) (approving use of 50-50 prior odds in case involving a severely retarded woman who became pregnant while institutionalized in a facility where only a small number of men had access to her). Cf. Allen et al., *supra* note 72, at 163 (criticizing 50-50 assumption and other “assessment of the odds of guilt or liability before the receipt of . . . evidence”).

¹⁵⁵ See, e.g., *In re the Paternity of M.J.B.*, 425 N.W.2d 404, 409-11 (Wis. 1988) (rejecting 50-50 assumption as insensitive to particular cases and allowing both parties to call expert witnesses to testify to their own and challenge the other’s prior odds assumption).

¹⁵⁶ See, e.g., *State v. Spann*, 617 A.2d 247, 250 (N.J. 1993) (holding that expert witness testifying to probabilities in criminal case must list and explain effect of different prior probabilities); *M. v. Marvin S.*, 656 N.Y.S.2d 802, 809 (N.Y. Fam. Ct. 1997) (affirming paternity judgment based on testimony by expert who “utilized the ‘chart approach’” to show “effect of a range of Prior Probability values on the actual genetic test results”); *Plemel v. Walter*, 735 P.2d 1209, 1217-20 (Or. 1987) (requiring paternity expert to use “chart approach” to identify outcomes based on prior odds estimates ranging from 0 to 100 percent). See also Kaye et al., *New Wigmore*, *supra* note 104, at 488-91 (endorsing chart approach); Dale A. Nance & Scott B. Morris, *Juror Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random-Match Probability*, 34 J. Leg. Stud. 395, 403, 429-436 (2005) (presenting results of empirical study showing that “chart format” reduces error by mock jurors using Bayesian analysis).

¹⁵⁷ See Kaye, *supra* note 96, at 191-92 (describing courts’ increasing impatience with challenges to forensic DNA probabilities).

¹⁵⁸ See *supra* notes 68-75 and accompanying text. See also Tribe, *supra* note 67, at 1365 (predicting that mathematical analysis, even if based on accurate calculations of probabilities and statistical odds, would make little headway in courts).

¹⁵⁹ See *supra* notes 140-143, 152-156 and accompanying text.

As for data-mining, it is true that genetic techniques are more advanced and precise than those available to isolate and determine the frequency of such non-forensic traits as shirt color and preferences among cigarette brands. Recall, however, that early DNA analysis struggled to distinguish one genetic locus from another and objectively determine the length of the tandem repeats at each locus, providing strong bases for objection in court. It was only through the medium of the adversary system that those objections motivated the invention of more exacting—if still imperfect—probes and other analyses.¹⁶⁰ Similar objections are forcing fingerprint analysis to acknowledge its status and improve its performance as way of systematically aggregating the effect of multiple, individually inconclusive matches.¹⁶¹ The next section discusses advances in a wide array of data-mining techniques that adversarial scrutiny might similarly domesticate to permit the reliable aggregation of probabilities associated with a host of non-forensic identifying traits.

C. Modern Data Mining and the Broad Availability of Frequency Information

Start with an easy case. Suppose evidence indicates that the white adult male perpetrator of a rape was left-handed (as are about 9% of the white male population), shorter than the 5'3" victim (as are 5% of that population), green-eyed (13% of the population), and balding (27% of the population). The donor of blood secreted into semen found in the rape victim was A negative (6% of the relevant population), and the donor was a "secretor" (80% of the population). Police receive an anonymous call from someone claiming to know nothing about any particular victim who says that an acquaintance—who is not described except to say he is a white man currently standing alone on the northeast corner of Fifth and Main Streets—told her he just raped a woman. Police arrive at that corner fifteen seconds later and stop a white adult male standing there alone. He turns out to be left-handed, green-eyed, balding and an A-negative secretor. Assuming the information from the victim and anonymous caller are accurate, and that left-handedness, eye color, height, blood type and secretor status are independent traits among white adult males, the probability that this white adult male encountered "at random"—i.e., for reasons independent

¹⁶⁰ See supra notes 137-139 and accompanying text.

¹⁶¹ See supra note 78 and accompanying text.

of the traits being matched—would be green-eyed, 5'3" or shorter, left handed, bald, A-negative and a secretor is estimated by multiplying the frequency of the identifying traits: $.09 \times .13 \times .05 \times .27 \times .06 \times .8$, or just under 8 in 1,000,000.

Because information on these traits has been collected for years, they avoid much of the criticism leveled against the mathematics instructor's testimony at the 1964 *Collins* trial for having no evidentiary basis for, and simply making up, frequencies for traits such as yellow sedans, pony-tailed white blonde women and bearded black men in Los Angeles.¹⁶² DNA initially faced the same complaint: law enforcement collected too few samples to permit reliable estimates of the frequency of particular alleles in the population.¹⁶³ Forty years after *Collins*, however, criticism of data mining runs in the opposite direction: we collect too much.¹⁶⁴ Mountains of collected information now allow us to construct accurate estimates of the frequency of the identifying traits in all three examples above—the rape case, DNA and the *Collins* case—either in the population at large or (at least for the *Collins* factors) in particular neighborhoods or at particular intersections. A simple internet search can produce frequency statistics for any number of reference populations of characteristics such as left-handedness, male and female factor baldness, cigarette smoking, car makes and models, eye color, size, weight, and more.¹⁶⁵ The cell phone, credit card, debt collection, electronic mapping, insurance, internet sales, marketing, medical, private investigation, security, search engine, social networking, and tracking (GPS) industries have massive amounts of data about many human characteristics, behaviors, patterns of dress, and other preferences, tastes and habits, much of which can be segmented by state, city and postal code.¹⁶⁶

¹⁶² See Tribe, *supra* note 67, at 1335 (criticizing prosecution's probabilistic testimony in *Collins* case as devoid of empirical support for probabilities used).

¹⁶³ See Kaye, *supra* note 96, at 88 (citing early criticism of DNA evidence based on inadequate DNA data sets using small samples from FBI recruits).

¹⁶⁴ For civil liberties criticisms of government collection of DNA samples, see, e.g., Ben Prottess, *The DNA Debacle: How the Federal Government Botched the DNA Backlog Crisis*, ProPublica, May 5, 2009, available at <http://www.propublica.org/special/dna-tracker-crime-labs-and-their-dna-backlog-724>; Michael T. Risher, *Racial Disparities in Databanking of DNA Profiles*, GeneWatch, <http://www.councilforresponsiblegenetics.org/GeneWatch/GeneWatchPage.aspx?pageId=204>; Jeffrey Rosen, *Genetic Surveillance for All*, Slate, March 17, 2009, <http://www.slate.com/id/2213958/pagenum/all/>.

¹⁶⁵ See, e.g., *supra* note 100.

¹⁶⁶ *United States v. Jones*, 565 U.S. ___, ___ (2012) (Alito, J., concurring) (noting that "automatic toll collection systems create a precise record of the movements of motorists," whose cars often have "devices that permit a central station to ascertain the car's

Law enforcement agencies themselves diligently collect information capable of revealing the frequency of human traits.¹⁶⁷ Police in London and New York routinely use thousands of cameras at strategic locations to monitor pedestrian and automobile traffic. The resulting photographs and CCTV videotapes are instantly relayed to central data bases and are of sufficiently high resolution to distinguish faces and the color, make and design of clothing and automobiles.¹⁶⁸ Many private stores, malls, business and neighborhood associations, transportation hubs, and universities videotape their own public spaces, enabling frequency data to be minutely segmented by location, average age of individuals and other criteria.¹⁶⁹ Oakland, California police officers use body-mounted personal video cameras to record crimes, arrests and traffic-stops.¹⁷⁰ Hundreds of fixed and portable police cameras in New York and elsewhere are sufficiently discerning to “photograph [automobile] license plates at the rate of hundreds per minute and then convert those images to data—letters and numbers—that” computers then compare “to a so-called ‘hot list’ of information on such things as stolen vehicles and other violations.”¹⁷¹ Law-enforcement officials have long mined traveler data to construct drug courier profiles, such as methods of paying for

location,” while “wireless carriers [can now] track and record the location of users [of] more than 322 million wireless devices” in the U.S.); *id.* at __ (Sotomayor, J., concurring in the judgment) (noting frequency with which people now disclose phone numbers, URLs and e-mail addresses and “books, groceries, and medications” purchases to cellular and Internet service providers and online retailers). For a small sampling of the burgeoning literature, see Krzysztof J. Cios et al., *Data Mining: A Knowledge Discovery Approach* (2007); Zdravko Markov & Daniel T. Larose, *Data Mining the Web: Uncovering Patterns in Web Content, Structure, and Usage* (2007); Robert Nisbet et al., *Handbook of Statistical Analysis and Data Mining Applications* (2009).

¹⁶⁷ See, e.g., Colleen McCue, *Data Mining and Predictive Analysis: Intelligence Gathering and Crime Analysis* (2006); Monica C. Holmes et al., *Data Mining and Expert Systems in Law Enforcement Agencies*, 8 *Issues Info. Syst.* 329 (2007); H. Chen et al., *Crime Data Mining: A General Framework and Some Examples*, 4 *Computer* 50 (2004).

¹⁶⁸ See, e.g., Al Baker, *Camera Scans of Car Plates Are Reshaping Police Inquiries*, *N.Y. Times*, April 12, 2011, at A17 (discussing London’s and New York City’s systems of security cameras linked to police coordination centers).

¹⁶⁹ See, e.g., Duncan Graham-Rowe, *Smart Statistics Keep Eye on CCTV*, *New Scientist*, Nov. 13, 2003, at <http://www.newscientist.com/article/dn4380-smart-statistics-keep-eye-on-cctv.html> (estimating as of a decade ago that there were 25 million CCTV cameras worldwide).

¹⁷⁰ Erica Goode, *Video, a New Tool for the Police, Poses New Legal Issues, Too*, *N.Y. Times*, Oct. 12, 2011, at A14.

¹⁷¹ Al Baker, *License Plate Reading Cameras Gain Legal Backing*, *N.Y. Times*, April 14, 2011, <http://cityroom.blogs.nytimes.com/2011/04/14/license-plate-reading-cameras-gain-legal-backing/>; Baker, *Camera*, *supra* note 168. Consider a recent example in the news:

[F]or all Mr. Zeledon’s evasiveness, the key to his arrest on murder charges in 2009 came days before the killing even occurred—as he was driving his car.

[P]hotos of [Zeldon’s] 2004 red Nissan Sentra with Connecticut license plates were captured and preserved by a network of police cameras and computers. Mr. Zeledon then became the prime suspect in the fatal stabbing of Andy Herrera, 28, on Jan. 19, 2009. . . .

. . . The clues were collected by a detective at 1 Police Plaza who had pulled them from databases and flashed them on a screen—making a map with the suspect’s photo at its center and a web of white lines connecting him to all his known locations, movements and associations.

Baker, *License Plate*, *supra*. See also *United States v. Jones*, 565 U.S. at __ (Sotomayor, J., concurring in the judgment) (“GPS monitoring [by police and others] generates a precise, comprehensive record of [individuals’] public movements that reflects a wealth of detail about [their] . . . associations.”).

airline tickets, “source” and “destination” cities, duration of stay, baggage and modes of dress. Now, they also trawl airline records, checkpoints and photographs for passengers who meet the predefined criteria, and resulting evidence is often admitted in court.¹⁷²

Of course, these activities raise serious civil liberties issues and are subject to proprietary, privacy and security limitations on information sharing.¹⁷³ Most of these concerns, however, relate to the use of information to single out specific individuals for unwanted intrusions, such as a sales pitch or search and arrest. Fewer concerns surround the pooling of information solely to identify frequencies of particular traits and behaviors, such as the proportion of adult male pedestrians aged twenty-five to fifty in a particular part of town with mustaches who wear brown sweatshirts and sprint rather than walk along the sidewalk at a given location between 3:00 and 5:00 pm in April.¹⁷⁴ Given the burgeoning use of artificial intelligence technology to sort data by features, these traits need not even be counted by humans, though they can be.¹⁷⁵ In conjunction with other information markets, demand by police, forensic laboratories, prosecutors, defense lawyers¹⁷⁶ and courts surely could induce entrepreneurs and agencies to mine existing and create new data sources to generate frequencies of detectible human attributes in given reference populations. Indeed, Target Corp., the national retailer, has created offices and laboratories in

¹⁷² See, e.g., *United States v. Sokolow*, 490 U.S. 1 (1989) (affirming reasonableness of search premised on defendant’s match to a drug courier profile); Mark J. Kadish, *The Drug Courier Profile: In Planes, Trains, and Automobiles; and Now in the Jury Box*, 46 Am. U.L. Rev. 747,754-55 (1997).

¹⁷³ See, e.g., *United States v. Jones*, 565 U.S. at __ (holding that, absent a warrant, officers’ entry onto private property to attach a GPS tracking device to Jones’ car and tracing of his movements for four weeks, constituted an unconstitutional search and seizure); Daniel J. Solove, *Data Mining and the Security-Liberty Debate*, 75 U. Chi. L. Rev. 343, 356 (2008).

¹⁷⁴ See, e.g., Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g (2010); Health Insurance Portability and Accountability Act (HIPAA), 42 U.S.C. § 201 (2010) (both barring release of data with individual identifiers but permitting release of aggregate data from which individual identifiers have been removed, for research and other purposes).

¹⁷⁵ See *International Journal on Document Analysis and Recognition*, which “includes contributions dealing with computer recognition of characters, symbols, text, lines, graphics, images, handwriting, signatures, as well as automatic analyses of the overall physical and logical structures of documents, with the ultimate objective of a high-level understanding of their semantic content.” <http://www.springer.com/computer/image+processing/journal/10032>.

¹⁷⁶ Prosecutors and defense lawyers are already taking advantage of related possibilities. See *Fine Tuning Technology in Today’s Economy*, 28th Ann. Conf. of the Nat’l Ass’n for Just. Info. Syst., Nashville Tennessee, Sept. 15-17, 2009, <http://events.linkedin.com/2009-NAJIS-28th-Annual-Conference/pub/61963>; *Forensics 411: How to Find Byte Marks*, Ann. Innocence Project of Minn. Program, Minneapolis, April 11, 2010, http://www.ipmn.org/forensics_411_2010.php. Topics at the latter legal-training program included “recovering lost clusters of the digital DNA files . . . from a crime lab’s computers”; “Forensic Aspects of Cell Phones, IM, Mobile Technologies, and Social Networking”; “Forensic Document Examination”; “Video Evidence In The Courtroom” and an inventory of retail services for “forensic video, audio and image analysis” and “latent print processing”).

Minneapolis and Las Vegas that offer just these services, including “sting trailers,” “mobile and stationary surveillance assistance” and “intelligence analysis.”¹⁷⁷

Objections surely will arise based on a lack of sufficient acuity or standards to distinguish videotaped, photographed or otherwise recorded traits reliably and determine their frequency in relevant populations. Who counts as an adult, rather than a mature-looking teenager? Where do the lines between brown, russet and red sweatshirts lie? What is the difference between a beard and a week’s worth of stubble? But similar issues already arise in court when surveillance footage is used for identification purposes, and there is no reason in principle why they are any less amenable to adversarial testing and increasing acuity and standardization than the examination of radiographs of chemically probed DNA samples to distinguish one allele from another based on length and thickness.¹⁷⁸ If there is a will to aggregate the probabilities of thousands of human attributes that can point to or away from particular criminal suspects, there is the way to estimate them reliably.

D. Aggregative Use of Non-Exclusionary Non-Matches Outside of Court

Our discussion thus far has focused on using aggregative analysis to reveal a reasonable doubt *in court*. As non-exclusionary non-matches in the DeLuna case reveal, however, police and prosecutors are as at least as less guilty of underusing aggregative analysis as the courts.¹⁷⁹ Unlike criminal trials, police investigations are not legally constrained from using statistical techniques. Yet, police often do not notice, much less capture,¹⁸⁰ the raw materials for such analysis: the full array of potential identifying markers associated with a crime that may or may not match later-identified suspects. For reasons we develop

¹⁷⁷ Target Corp., Intelligence: Gathered, Analyzed, Disseminated & Managed, Presentation at Nat’l Ass’n for Justice Info. Syst. Conf., Nashville, Sept. 15-17, 2009, at 1, <http://www.najis.org/Presentations/2009%20Presentations/NAJIS%20Handout%20Sept%2015-17%202009%5B1%5D%20target.pdf>; see Target Corp. Law Enforcement Partnerships: Forensic Overview, <http://pressroom.target.com/pr/news/forensics-support.aspx?ncid=24622> (noting that Target’s forensic labs in Minneapolis and Las Vegas received American Society of Crime Lab Directors’ “accreditation for digital evidence (video analysis, audio analysis and image analysis)”).

¹⁷⁸ See, e.g., *United States v. Shabazz* 564 F.3d 280, 286 (3d Cir. 2009) (admitting conflicting testimony on whether defendant appeared in surveillance footage); *Washington v. State*, 961 A.2d 1110, 1114 (Md. 2008) (treating admission of compiled images from surveillance videotapes as non-harmless error, given lack of authentication); *supra* notes 122-161 and accompanying text.

¹⁷⁹ See *supra* notes 26 and accompanying text.

¹⁸⁰ For example, good police practice calls for investigators to lay down a fine-meshed grid at crime scenes, then videotape and meticulously examine the contents of each cell. See *infra* note 436.

below, police (as well as everyone else) are congenitally more interested before the fact in “big” evidence, such as confessions and eyewitness identifications, than “small” matches. Then, after the fact, police have no incentive to attend to or disclose what have turned out to be “small” *non*-matches.¹⁸¹ Notwithstanding our own, and the academic literature’s, fascination with the possibility of aggregative analysis in court, the uses and benefits of those techniques outside of court are even more important to factor into our analysis below.

V. COGNITIVE, STRUCTURAL AND LEGAL IMPEDIMENTS TO USING NON-EXCLUSIONARY NON-MATCHES

Above, we posit a burglary-murder case in which police base a strong case of guilt against a suspect on an eyewitness identification and a combination of the female victim’s missing stocking tops and discovery of different stocking tops in the defendant’s possession. We then show how probabilities associated with a series of non-exclusionary non-matches—evidence that the fatal blows were administered left-handed, though the defendant is right-handed; a cigarette butt found in the victim’s foyer, though neither she nor the defendant smokes; and features of the defendant that do not match descriptions of a man seen at the crime scene—could aggregate to a reasonable doubt that would likely escape notice absent systematic aggregation.¹⁸² In this part, we explain the prediction that, absent systematic aggregation, powerful cognitive, structural and legal forces lead actors in the criminal justice process to undervalue “small” non-match evidence and risk convicting the innocent.

A. Cognitive Resistance

1. *Heuristic Economization.* — Human heuristic biases—unconscious cognitive tendencies to oversimplify the evaluation of uncertain probabilities in all facets of decision making—help explain the overvaluation of “big” matches and undervaluation of “small” non-matches.¹⁸³ Starting with the work of

¹⁸¹ See *infra* notes 285-301, 341-352 and accompanying text.

¹⁸² See *supra* notes 37-42 and accompanying text.

¹⁸³ See, e.g., Timur Kuran & Cass R. Sunstein, Availability Cascades and Risk Regulation, 51 *Stan. L. Rev.* 683, 704 (1999) (“Because our cognitive limitations preclude us from thinking deeply about more than a small fraction of the issues that bear on our values, behavior, and welfare, we rely on mental shortcuts that leave us misinformed in many contexts, even seriously wrong.”); Paul Slovic et al., Facts Versus Fears: Understanding Perceived Risk, in Daniel Kahneman et al., eds., *Judgment Under*

Nobel Laureate Daniel Kahneman, a rich literature documents the human tendency to reach judgments that seem intuitively correct but are logically and empirically false because they screen out categories of information that basic physical rules of the universe make crucial to the achievement of accurate judgments.¹⁸⁴

2. *The Representativeness Bias.* — Consider the representativeness bias, which strongly predisposes human decision makers inaccurately to estimate the probability that individuals in category “A” (defined by one or more personal traits) also have characteristic “B” (defined by a different trait or set of them) by asking how often individuals with characteristic B also have characteristic A.¹⁸⁵ Imagine a police detective assessing the chance that the person who committed a robbery (characteristic A) for which there are many possible suspects has a prior record of robbery convictions (characteristic B). Because characteristic B (being a robber) is “representative” of, or resembles, characteristic A (committing a recent robbery), the intuitive assessment of the probability that a “known” prior robber committed a recent robbery tends to be very high. In fact, however, the probabilistically correct answer depends not only on prior robber’s resemblance of the current robber but also on the base rate of people in the suspect pool with characteristic B (having a prior robbery conviction). Although a past robber probably is more likely than the average person with no prior record of robbery to commit a new robbery,¹⁸⁶ it is not necessarily true that a recent robbery was probably committed by someone with a prior robbery conviction. Because the proportion of prior robbery convicts in the pool of all possible suspects is likely to be fairly low—most people with access to the crime scene probably have no criminal

Uncertainty: Heuristics and Biases 464-65 (1982) (arguing that heuristic biases “reduce difficult mental tasks to simpler ones [and] are valid in some circumstances, [but] in others they lead to large and persistent biases with serious implications”).

¹⁸⁴ See, e.g., Daniel Kahneman, *Thinking, Fast and Slow* 7-8 (2011); Daniel Kahneman & Amos Tversky, *Judgment and Decision Making: An Interdisciplinary Reader* 35 (2005).

¹⁸⁵ See Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 *Cog. Psych.* 430 (1972) (“A person who follows this heuristic evaluates the probability of an uncertain event, or a sample, by the degree to which it is: (i) similar in essential properties to its parent population; and (ii) reflects the salient feature of the process by which it was generated.”). For example, presented with a description of an individual with traits stereotypically associated with librarians and asked to assess whether it is more likely that the person is a farmer, salesman, physician or librarian, subjects consistently, but mistakenly, pick librarian. Even if only a small fraction of salesmen are “bookish” while nearly all librarians are, a “bookish” person is more likely to be a salesman than a librarian, because there so many more salesmen than librarians. See Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 *Science* 1124, 1124 (1974).

¹⁸⁶ See Lempert et al., *supra* note 81, at 483 n.251 (discussing recidivism rates for robbery).

record—it may be more likely that a recent robbery was committed by a first-timer than by a repeat-offender. The representativeness bias thus disposes observers—police, prosecutors and jurors included¹⁸⁷—to focus on the fact that a convict-suspect is more likely to be the robber than any *one* of the potential suspects without a criminal record and to ignore the more important fact that the culprit more likely is a member of the *group* of potential suspects who do not have a prior record. An intuitively satisfying but probabilistically risky effort to search for more evidence to implicate the prior convict—or even his arrest—may ensue, rather than the additional investigation of *all* suspects that could well be called for. More generally, and regardless of occupation and expertise, human decision makers seem to be hardwired to use only resemblance—the seemingly more “individualized” or “personalized” evidence—and not base rate information when both have valuable information to contribute.¹⁸⁸

Bayes’ Theorem helps to formalize the mistake the representativeness bias impels. Adapting one of Kahneman’s famous experiments,¹⁸⁹ suppose we know that X, an unidentified bank robber, exhibits behaviors associated with the stereotype of a radical bookseller (e.g., is evidently a fan of obscure utopian novels like those found in the getaway car) and we want to know how likely it is that X is a bookseller, in order to decide whether to expend resources investigating members of that occupation. Bayes’ Theorem provides the correct equation for answering this question: the prior odds that an individual in the suspect pool is a bookseller (B)—stated as the ratio of booksellers to non-booksellers in the relevant population—times the likelihood that X is a bookseller given that X has characteristics stereotypically associated with radical booksellers. The latter likelihood is also a ratio: the probability (P) that the evidence (E) would exist if the culprit is a bookseller (B) divided by the probability the evidence would exist if culprit is *not* a

¹⁸⁷ See Andrew E. Taslitz, *Police Are People Too: Cognitive Obstacles To, And Opportunities For, Police Getting the Individualized Suspicion Judgment Right*, 8 Ohio St. J. Crim. Law 7, 43-44 (2010) (discussing susceptibility of police to cognitive biases).

¹⁸⁸ See, e.g., W. Casscells et al., *Interpretation by Physicians of Clinical Laboratory Results*, 299 N. Eng. J. Med. 999, 999-1001 (1978) (finding that Harvard Medical School faculty tended to neglect base rates when estimating the effect of a specified rate of false positive diagnoses for particular disease); I. Dror & D. Charlton, *Why Experts Make Errors*, 56 J. Forensic Identification 600, 612 (2006); Tversky & Kahneman, *supra* note 185, at 1130 (“The reliance on heuristics and the prevalence of biases are not restricted to laymen. Experienced researchers are also prone to [them.]”); Richard H. Thaler, *Illusions and Mirages in Public Policy*, in *Judgment and Decision Making*, *supra* note 184, at 166 (“Cognitive illusions influence representatives, senators, presidents—even so-called experts are not immune.”).

¹⁸⁹ See *supra* note 185.

bookseller.¹⁹⁰ More simply, the likelihood ratio is the probability that booksellers behave in the way X did divided by the probability that non-booksellers behave that way. So:

$$\text{Prior Odds} \times \text{Likelihood Ratio} = \text{Subsequent Odds}$$

$$(B's \text{ in suspect pool} / \text{non-}B's \text{ in suspect pool}) \times (P E(B) / P E(\text{not } B)) = \text{Subsequent Odds}$$

Assume that booksellers comprise 5 percent of the suspect population, so the prior odds are 5 to 95 (1 to 19) that X is a bookseller. Even if 97 percent of booksellers tend to exhibit the behavior X exhibited, and only 6 percent of the rest of the population do, there still is a less than even chance (49 percent) that X is a bookseller:

$$1/19 \times .97/.06 = .97/1.14 \rightarrow 46\%$$

Although X probably is *not* a bookseller, the representativeness bias will lead most observers to the opposite conclusion—round up (or, at least, investigate) the nearest booksellers! The representativeness shortcut is to focus on the probative value of X's bookish trait—that is, how much greater the likelihood-ratio *numerator* is than the *denominator* (.97 vs. .06)—and ignore the effect of the *prior odds*.

Consider next the effect of the representativeness bias in a criminal case in which the evidence against the defendant is an eyewitness identification as to which there is small chance of a mistake, say two percent. The probability of guilt is the prior odds that the defendant is guilty *times* the likelihood ratio associated with the identification. The prior odds are a function of the number of suspects. If the defendant is one of 5 suspects, the prior odds of guilt are 1 to 4. If the defendant is one of 5000 suspects, the prior odds are 1 to 4,999. Here, the representativeness bias may lead police, prosecutors and jurors to treat the identification as equally powerful, whether there are 101 possible suspects or only three, even though in the former case, the probability of guilt is at best only 33 percent ($1/100 \times 1/.02 = 1/2 \rightarrow 33\%$), while in the latter case, the probability of guilt approaches 96 percent ($1/2 \times 1/.02 = 1/.04 \rightarrow 96\%$).

This last example helps explain why study subjects tend to value eyewitness identifications and confessions more than circumstantial evidence, say, a ballistic match, that is as strong or stronger.¹⁹¹

¹⁹⁰ See supra notes 40-42 and accompanying text.

Unlike an individual singled out as the culprit by an eyewitness or by the suspect's own admissions, a bunch of striations on a hunk of metal do not resemble our idea of a criminal. The overvaluing of "personalized" evidence relative to even strong forensic evidence holds true even when experimental subjects are presented with information quantifying the prior odds of guilt or liability, the probabilities associated with the forensic or other evidence, and the probability that an eyewitness is mistaken given a documented history of perceptual mistakes in the same situation.¹⁹² Even when the possibility of inaccurately estimated subjective probabilities is removed, study subjects give more weight to "individualized" or "representative" eyewitness and confession evidence than to other, stronger evidence.¹⁹³

When presented with an eyewitness identification, confession or other personalized evidence that "represents" a guilty person, human decision makers intuitively anchor on the *likelihood ratio* associated with the evidence and ignore the *prior odds*, or effect of the number of potential suspects. But when the evidence is overtly probabilistic, the bias does not come into play, and intuitive decision makers are likely to be more attentive to whether there are other suspects who might match the evidence. The representativeness bias will particularly disadvantage the defense relative to the prosecution, obscuring

¹⁹¹ See, e.g., Heller, *supra* note 18, at 244-45, 248-50 (discussing studies, including one in which "jurors overestimated the accuracy of eyewitness identifications by more than 500%" (citing John C. Brigham & Robert K. Bothwell, *The Ability of Prospective Jurors to Estimate the Accuracy of Eyewitness Identifications*, 7 *L. & Hum. Behav.* 19, 24 (1983)); Deanna L. Sykes & Joel T. Johnson, *Probabilistic Evidence Versus the Representation of an Event: The Curious Case of Mrs. Prob's Dog*, 21 *Basic & Applied Soc. Psychol.* 199, 208 (1999) (similar); Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 *J. Personality & Soc. Psych.* 739, 742 (1992).

¹⁹² In an influential experiment, Professor Wells presented mock jurors with two scenarios in a civil case involving a dog run over by a bus that could only have been operated by the Blue or Grey Bus Company, each with the same number of vehicles. In one case, a weigh station official logged in a "Blue Bus" on the road in question ten minutes before the dog was run over ten minutes down the road, but acknowledged that his records mistake Grey for Blue Buses 20 percent of the time. In the other case, tire track marks at the scene of the accident were found to match 80 percent of the Blue and only 20 percent of the Grey Buses. Although Wells designed both scenarios to create an equal, 80 percent probability of Blue Bus liability, judges and psychology and business students were, respectively, four, five and nine times more likely to find the Blue Bus Company liable in the eyewitness case than in the tire-track case. See Wells, *supra* note 191, at 741-44.

¹⁹³ See Heller, *supra* note 18, at 244, 255-58 (arguing that juror overvaluation of eyewitness over forensic testimony is due not to misvaluation of evidence but to a "psychological" tendency to ignore certain kinds of evidence). See also David L. Faigman & A.J. Baglioni, Jr., *Bayes' Theorem in the Trial Process: Instructing Jurors on the Value of Statistical Evidence*, 12 *L. & Hum. Behav.* 14 (1988) (discussing studies documenting "individuals' reluctance to use statistical information when making causal attributions"); Brian C. Smith et al., *Jurors' Use of Probabilistic Evidence*, 20 *L. & Hum. Behav.* 49, 52-54 (1996) (discussing studies showing persistence of heuristic mistakes even after jurors were trained in Bayesian analysis). But see *infra* notes 392-422 and accompanying text (discussing research indicating that graphic and other simplified representations of Bayesian analysis can substantially increase lay decision makers' use, understanding and accuracy in applying Bayesian reasoning).

reasonable doubt that exists, in the common situation in which the prosecution's case is mainly based on an identification or confession and the defendant's evidence is mainly based on a collection of only "small" non-matches.¹⁹⁴ The disadvantage likely will remain even if the non-matches undermine the personalized evidence itself—as when traits of a suspect identified by an eyewitness don't match the witness's initial description of the culprit or a confession includes details contrary to the known facts of the crime.

Adding to the problem, experimental studies show that human decision makers don't simply overvalue eyewitness identifications and confessions (by ignoring prior odds), while giving other evidence of guilt the *correct* weight. Human intuitions systematically give so-called non-"direct" or "circumstantial" evidence of guilt—including forensic evidence—less weight than Bayes' Theorem shows it deserves.¹⁹⁵ Professor Kevin Jon Heller's comprehensive review of the research exposed "an unsettling paradox": although circumstantial evidence of guilt "is far less likely to lead to a false conviction than direct evidence, jurors are . . . reluctant to use it to convict," to the point of risking "false acquittal."¹⁹⁶

The representativeness bias helps explain why "direct" evidence is overvalued relative to a correct Bayesian analysis but not why "circumstantial" evidence is undervalued. The next three sections discuss other cognitive biases that explain the latter effect and particularly the undervaluation of non-exclusionary non-matches.

3. *The simulation, confirmation and certainty biases.* — Professor Heller has identified three heuristic biases—simulation, confirmation and certainty—that dispose jurors to undervalue circumstantial

¹⁹⁴ See *supra* notes 6-26 and accompanying text and *infra* notes 270-276 and accompanying text (discussing Mathews/Hayes, DeLuna, and Mayfield cases).

¹⁹⁵ See Heller, *supra* note 18, at 251-52 (describing study involving a hypothetical murder case in which five groups of mock jurors invited to use blood-typing evidence to assess the probability of the defendant's guilt underestimated the probative value of the evidence relative to the Bayesian norm by 80-100 percent, with the disparity being "greatest when the evidence was the most incriminating" (citing Jane Goodman, *Jurors' Comprehension and Assessment of Probabilistic Evidence*, 16 *Am. J. Trial Advoc.* 361, 368-73 (1992))). See also Paul Bergman, *A Bunch of Circumstantial Evidence*, 30 *U.S.F. L. Rev.* 985, 986 (1996) (noting public misconceptions about the weakness of "circumstantial evidence"); Lisa Smith et al., *supra* note 25, at 410, 414 & nn.9-11, 15 (citing sources).

¹⁹⁶Heller, *supra* note 18, at 245.

evidence of guilt and overvalue identifications, confessions and other direct evidence. Although Heller's juxtaposition of undervalued "circumstantial" and overvalued "direct" evidence of guilt is different from our distinction between undervalued "small" non-match evidence of innocence and "big" evidence of guilt, including "circumstantial" DNA and fingerprints, as well as "direct" identifications and confessions, his analysis aids our argument and bears summary.

The simulation bias. — The simulation bias leads individuals who imagine a scenario in which X is true (e.g., that candidate X will beat candidate Y in an election) to believe thereafter that the probability of X is higher than she previously believed it to be.¹⁹⁷ This tendency holds even when subjects are presented with information that, from a Bayesian perspective, should make them realize that the pre- and post-simulation probabilities are the same.¹⁹⁸

Heller argues convincingly that eyewitness identification testimony and confessions trigger the simulation bias by providing a "high-coverage" narrative about what happened that automatically, if not always accurately, increases the intuitive decision maker's assessment of the probability that the scenario is true.¹⁹⁹ By contrast, "circumstantial" evidence of guilt, even fingerprint or DNA evidence that Bayesian analysis reveals to be stronger, has no simulation effect. It wears a counter-factual on its sleeve: the (albeit often quite small) possibility that someone else has the same web of lines in a part of their fingerprint or string of alleles in their DNA or has framed the defendant.²⁰⁰ Adding to the disparity, direct evidence in the form of an eyewitness identification or confession is "vivid": it is "representational" in appearing to present a single reality, "narrative" in that it comes in the especially accessible form of a relatively coherent story, "univocal" because it points in a single direction, and claims to be

¹⁹⁷ See John S. Carroll, The Effect of Imagining an Event on Expectations for the Event: An Interpretation in Terms of the Availability Heuristic, 14 J. Exper. Soc. Psychol. 88, 90-92 (1978) (describing research indicating that imagining a scenario increases subjects' assessment of its likelihood); Heller, supra note 18, at 260-61; Daniel Kahneman & Amos Tversky, The Simulation Heuristic, in Kahneman et al. eds., supra note 197, at 201. See also Philip Broemer, Ease of Imagination Moderates Reactions to Differently Framed Health Messages, 34 Eur. J. Soc. Psychol. 103, 115-16 (2004) (showing that anti-smoking and other public-health messages are more likely to affect behavior when they force viewers to imagine particular behaviors and outcomes). The hindsight bias is a version of the simulation bias, where experiencing an actual event makes it seem more foreseeable than it was. See, e.g., Harmut Blank et al., Hindsight Bias: On Being Wise After the Event, 25 Soc. Cognition 1 (2007).

¹⁹⁸ See Heller, supra note 18, at 260-61 (citations omitted).

¹⁹⁹ See id. at 259. "Coverage" refers to the proportion of the overall event that the evidence portrays or "covers."

²⁰⁰ See id. at 265.

“unconditional.”²⁰¹ By contrast, circumstantial evidence is “pallid,”²⁰² i.e., “abstract”—or, one might say, class-based—because it reports what is true of categories of phenomena, “rhetorical” in that it comes in the harder-to-digest form of an argument (if X, then probably Y), “polyvocal” because it suggests multiple possibilities, and “probabilistic.”²⁰³

The vividly simulating effect of eyewitness statements and confessions is likely to sway police and prosecutors in deciding whom to arrest and charge, especially when that evidence is the only high-coverage scenario before them. Once they are exposed to such a scenario that reasonably seems to explain what happened, they will tend to “cease the simulation process and fail to consider alternative scenarios that imply a different outcome.”²⁰⁴ The effect may be somewhat mitigated at trial because defense counsel and, during deliberations, other jurors can offer alternative theories about what happened and emphasize the regret jurors should feel if they convict an innocent person—two mechanisms that can diminish the simulation effect.²⁰⁵ But insofar as the alternative theory is based either on circumstantial evidence or reasonable doubt—i.e., is pallid, abstract, rhetorical, polyvocal and probabilistic—it will not supply a compelling counter-narrative of what happened or get any of the extra “simulation” or “representativeness” bounce that eyewitness and confession evidence gets.²⁰⁶ Additionally, the tendency of jurors to overvalue the reliability of what people say, especially about themselves, gives eyewitness and confession evidence an additional systematic advantage over “circumstantial” evidence.²⁰⁷

²⁰¹ See *id.* at 264, 269.

²⁰² See *id.* at 269.

²⁰³ See *id.* at 264, 269. See also Richard Greenstein, *Determining Facts: The Myth of Direct Evidence*, 45 *Hous. L. Rev.* 1801, 1815, 1829-30 (2008-09) (offering similar explanation of the “linguistic trick” that leads people to treat a witness’s “nam[ing of] the ‘fact of consequence’ directly” as providing more immediate access to the truth than circumstantial evidence, which “names a different fact” that “is connected to the fact of consequence [only] by inferential steps”).

²⁰⁴ Edward R. Hirt & Keith D. Markman, *Multiple Explanation: A Consider-an-Alternative Strategy for Debiasing Judgments*, 69 *J. Personality & Soc. Psychol.* 1069, 1084 (1995).

²⁰⁵ See Heller, *supra* note 18, at 281 (noting that “priming”—calling to mind alternative scenarios—“is determined by two factors: whether the structure of the decision-making task encourages the consideration of alternative scenarios; and. . . involves ‘negative affect’ such as fear or regret.”).

²⁰⁶ See *id.* at 292, 294 (arguing that reasonable doubt and circumstantial evidence defenses deprive jurors of the sorts of evidence that are most capable of combatting simulation bias).

²⁰⁷ See *id.* at 285-86 (describing experimental evidence of “truth bias,” a tendency to believe another’s autobiographical statements regardless of their truthfulness, and “narrative transportation,” which leads listeners to accord more weight to the credibility of statements in the form of stories; because of both tendencies, jurors barely outperform chance when trying to detect whether a witness is lying or mistaken and become even less effective when told to be suspicious of particular categories of witnesses). See also Melanie C. Green & Timothy C. Brock, *The Role of Transportation in the Persuasiveness of Public*

The confirmation bias. — Even if police, prosecutors and jurors recognize that the scenario provided by an eyewitness or confessing suspect may be mistaken or fabricated, the “belief-perseverance” aspect of the confirmation bias still may deter them from considering alternative theories. Once humans adopt a theory, they tend to search for and give excessive weight to evidence that *confirms* it and to discount new evidence or interpretations of existing evidence that undermine it.²⁰⁸ Barbara O’Brien asked subjects to read a hypothetical criminal case file. Partway through reading the file, half of the participants were prompted to specify the person they believed committed the crime. The other participants identified the perpetrator only at the end. O’Brien found that the former subjects remembered more facts consistent with the guilt of their identified suspect than the latter subjects, picked more lines of investigation focused on that suspect and interpreted ambiguous evidence to be more consistent with that suspect’s guilt.²⁰⁹

Even when an investigator is driven to find every bit of evidence she can, the confirmation bias disposes her to organize the search based not on how much evidence she can find that points to all possible suspects but on how much she can find that points to the particular suspect she initially identified.²¹⁰ Once the representativeness and simulation biases lead investigators or jurors to overvalue an eyewitness’s or a confessing suspect’s narrative about what happened, therefore, individuals’ tendency to “adhere to their beliefs when the original evidential basis of the beliefs is shown to be flimsy, false, or

Narratives, 79 *J. Personality & Soc. Psychol.* 701, 719 (2000) (“[T]ransportation is unlikely in response to even very good rhetoric [but] can readily occur in response to a moderately compelling narrative.”); Heller, *supra* note 18, at 287-88 (“[When a hearer becomes immersed in a story,] all of her mental systems—attentive, imagistic, emotive—converge on its events, [and] her ability to think critically about the narrative is reduced, making her more likely to believe that it is authentic.”).

²⁰⁸ See, e.g., Karl Ask & Pär Anders Granhag, Motivational Sources of Confirmation Bias in Criminal Investigations: The Need for Cognitive Closure, 2 *J. Investigative Psych. & Offender Profiling* 43, 46 (2005); Alafair S. Burke, Improving Prosecutorial Decision Making: Some Lessons of Cognitive Science, 47 *Wm. & Mary L. Rev.* 1587, 1594-601 (2006); Alafair S. Burke, Neutralizing Cognitive Bias, 2 *N.Y.U. J.L. & Lib.* 512, 517 (2007); Keith Findley & Michael Scott, The Multiple Dimensions of Tunnel Vision in Criminal Cases, 2006 *Wis. L. Rev.*, 291, 309-17 (2006); Paul C. Giannelli, Confirmation Bias, 22 *Crim. Just.* 60, 61 (2007); Charles G. Lord et al., Considering the Opposite: A Corrective Strategy for Social Judgment, 47 *J. Personality & Soc. Psych.* 1231, 1232-40 (1984); Mark Snyder & William B. Swann, Jr., Hypothesis-Testing Processes in Social Interaction, 36 *J. Personality & Soc. Psychol.* 1202, 1205 (1978) (finding that simply phrasing an inquiry as whether a person is an “extrovert” or instead whether she is an “introvert” affects the conclusions reached).

²⁰⁹ Barbara O’Brien, Prime Suspect: An Examination of Factors that Aggravate and Counteract Confirmation Bias in Criminal Investigations, 15 *Psych., Pub. Pol. & L.* 315, 328-29 (2009).

²¹⁰ See *id.* at __; Jonep Sonnemans & Frans van Dijk, Errors in Judicial Decisions: Experimental Results 27 (Jan. 12, 2011), <http://jleo.oxfordjournals.org/content/early/2011/01/12/jleo.ewq019.abstract> (reporting results indicating that mock jurors stop searching for additional evidence even when rewarded for finding more and are disposed to convict before reaching a reasonable probability of guilt).

nonexistent”²¹¹ is likely to deter them from searching for alternative theories even after finding out that the witness may well be mistaken or untruthful.²¹² Belief-perseverance as to theories based on visibly discredited evidence is stronger, moreover, when the discredited evidence is part of a “coherent, causally related account in which a single or minimal correction has a significant impact on the construal of meaning.”²¹³ The bias is stronger when triggered by evidence in the form of even a weak narrative and is less strong when triggered by even powerful but isolated chunks of evidence.²¹⁴

Notice that every attribute of “circumstantial” evidence generally that keeps it from getting the artificial representativeness, simulation and confirmation bounces even more clearly weighs down the non-exclusionary non-match evidence of innocence that interests us here. Non-exclusionary non-matches are as pallid, abstract, rhetorical, non-transporting,²¹⁵ polyvocal and probabilistic as evidence can be. They are non-representational and non-narrative; often fall in a demeaned category of the “absence” or “negation” of evidence;²¹⁶ typically qualify as “small” evidence given their individually limited probative weight; by definition (“non-exclusionary”) come with obvious explanations for why they are present without bearing on guilt or innocence (e.g., that someone besides the culprit left the trace at the scene, or that small discrepancies in witnesses’ memories are inevitable); and gain strength only by being statistically aggregated with other evidence. The tendency of *all* circumstantial evidence of *guilt* to invite

²¹¹ Martin F. Davies, Belief Persistence After Evidential Discrediting, 33 J. Exper. Soc. Psych. 561, 562 (1997). See Craig A. Anderson, Abstract and Concrete Data in the Perseverance of Social Theories: When Weak Data Lead to Unshakeable Beliefs, 19 J. Exper. Soc. Psych. 93, 95 (1983).

²¹² See Heller, supra note 18, at 292-93 (“Once jurors conclude that the defendant is most likely guilty . . . , a confirmation bias sets in that limits their ability to recognize evidence inconsistent with that conclusion.”); Derek J. Koehler, Explanation, Imagination, and Confidence in Judgment, 110 Psychol. Bull. 499, 503 (1991) (presenting experimental evidence that once human decision makers gain confidence in a belief, “inertia sets in, which makes it more difficult to consider alternative hypotheses impartially”).

²¹³ Heller, supra note 18, at 288-89.

²¹⁴ Id. at 290 (“[B]elief-perseverance is strongest when the evidence supporting a belief is concrete and causally coherent. Circumstantial evidence is neither.”). See Sykes & Johnson, supra note 191, at 209-10 (noting that “comprehension of probabilistic information does not mandate a belief in the reality of a specific event,” while comprehension of a narrative of what happened does mandate that belief, so “belief engendered by an assertion about an event is more difficult to mutate than a belief based solely on statistical probabilities”).

²¹⁵ See supra note 207.

²¹⁶ See Stephen A. Saltzburg, A Special Aspect of Relevance: Countering Negative Inferences Associated with the *Absence* of Evidence, 66 Cal. L. Rev. 1011, 1019-21 (1978) (criticizing negative inferences from the absence of evidence).

convincing counter-theories and thus to carry the seeds of its own destruction is even more true of *non-exclusionary non-match evidence of innocence*.²¹⁷

The effect of the simulation, truth and confirmation biases may be formalized in the same Bayesian terms as we used to formalize the effect of representativeness. A Bayesian analysis of evidence of guilt multiplies the prior odds of guilt times the likelihood ratio, with the latter being defined as the probability that the evidence would exist if the suspect is guilty divided by the probability it would exist if the defendant is innocent. When triggered by narrative evidence such as a confession or eyewitness testimony identifying a culprit, the representativeness bias leads lay decision makers to ignore the *prior odds* variable in the equation. When triggered by the same narrative evidence, the simulation and confirmation biases seem to lead lay decision makers to anchor on the likelihood-ratio numerator (the probability that the incriminating evidence exists because the defendant is guilty) and to discourage consideration of the *denominator* (the probabilities associated with counter-scenarios under which the evidence exists though the defendant is innocent).

The certainty effect. — Another bias, the “certainty effect,”²¹⁸ helps explain why lay decision makers give less weight to circumstantial evidence compared than a proper Bayesian analysis requires. When facing risks, intuitive decision makers accord greater value than is rationally warranted to outcomes they believe are “certain,” i.e., that do or purport to eliminate the risk entirely. People “overweight outcomes that are considered certain, relative to outcomes which are merely probable”²¹⁹ and “greatly undervalue a reduction in the probability of a hazard in comparison to the complete elimination of that hazard.”²²⁰

²¹⁷ “[J]urors generally find it relatively easy to imagine [counter-theories] in a circumstantial case. The polyvocality of circumstantial evidence means that the prosecution’s own evidence is available for use in [building counter-theories], and strong priming normally ensures that jurors will pay close attention to any” competing evidence and counter-theory offered. Heller, *supra* note 18, at 299-300.

²¹⁸ Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision under Risk*, 47 *Econometrica* 263, 265-67 (1979).

²¹⁹ *Id.* at 265.

²²⁰ Daniel Kahneman & Amos Tversky, *Choices, Values, and Frames*, in *Choices, Values, Frames* 1, 9 (2000). See Kuran & Sunstein, *supra* note 183, at 707 (“Because people attach intrinsic value to certainty, their well-being improves more when the probability of an adverse effect drops from 1.0% to zero than when it drops from 2.1% to 1.0%.”); George F. Loewenstein et al.,

There is a built-in fudge factor, as well. Naïve decision makers sometimes treat the reduction to something just short of zero of even a very serious risk as if no risk remains, creating a false sense of security.²²¹ “Studies of insurance markets have found that we tend to ignore small risks until their probability passes a certain threshold, at which point we overspend wildly to prevent them.”²²² Additionally, differing ways of describing identical risks can nudge individuals into perceiving the situation as either presenting an excessively comforting zero risk—“pseudocertainty,” Tversky and Kahneman call it²²³—“or an excessively worrisome probability of harm. Asked to say whether they would volunteer to receive a vaccine that cuts in half the risk of contracting a serious disease expected to afflict 20 percent of the population, subjects are substantially less likely to do so than when asked to consider receiving a vaccine in the case of two equally serious and probable strains of a disease, each expected to afflict 10 percent of the population, that would reduce the probability of contracting one strain to zero but would have no effect on the other strain.”²²⁴ Although the risk reduction in both cases is the same—from 20 to 10 percent—the description of a treatment as reducing one of two equal risks to zero makes it more attractive than a treatment described as reducing the same overall risk by half.

Professor Heller hypothesizes that jurors faced with eyewitness testimony about what happened or a defendant’s confession are much more likely to treat it as establishing a “certainty” of guilt than even very strong circumstantial evidence that creates a much higher probability of guilt.²²⁵ Jurors may reach a certainty conclusion about eyewitness and confession evidence, Heller suggests, because such evidence establishes a 100 percent probability of guilt as long as the testimony is accurate, and because the simulation and confirmation biases and the ability to blame the witness, not themselves, if the testimony

Risk as Feelings, 127 *Psychol. Bull.* 267, 276 (2001) (showing that individuals will pay far more to reduce their risk of poisoning from five-in-ten-thousand to zero than from fifteen-in-ten-thousand to five-in-ten-thousand, though latter reduction is twice as large); Thaler, *supra* note 188, at 163-64 (similar).

²²¹ See Paul Slovic et al., Risk as Analysis and Risk as Feelings: Some Thoughts About Affect, Reason, Risk, and Rationality, 24 *Risk Analysis* 311, 318 (2004) (“[R]esponses to uncertain situations appear to have an all or none characteristic . . .”).

²²² Benjamin Wallace-Wells, Cass Sunstein Wants to Nudge Us, *N.Y. Times Mag.*, May 13, 2010, at 38 (discussing Richard H. Thaler & Cass R. Sunstein, *Nudge: Improving Decisions about Health, Wealth and Happiness* (2009)).

²²³ Amos Tversky & Daniel Kahneman, The Framing of Decisions and the Rationality of Choice, 211 *Science* 453, 455 (1981).

²²⁴ Paul Slovic, et al., *supra* note 183, at 480-81 (discussing vaccination study).

²²⁵ Heller, *supra* note 18, at 283-85.

is wrong gives jurors more confidence in the accuracy of the evidence than it deserves.²²⁶ By contrast, circumstantial evidence of guilt—no matter how reliable the testimony presenting it—always comes with some doubt.²²⁷ That doubt is magnified by the stress jurors experience when contemplating the possibility of an inaccurate verdict.²²⁸ As a result, Heller reasons, “[j]urors will dramatically underweight the ‘merely probable’ circumstantial case and dramatically overweight the ‘considered certain’ direct case—making the circumstantial case seem far more likely to result in a false conviction,” though the opposite often is true.²²⁹

There is a problem with Heller’s argument that he acknowledges but does not entirely solve: “the probative value of direct evidence is never 1.0 because . . . ‘the credibility of a witness always rests in part on circumstantial evidence,’” which, by hypothesis, always carries with it a possibility that the eyewitness or confession testimony is in error.²³⁰ If certainty arises only in the perceived absence of *any* overt possibility that the hazard will generate a harm,²³¹ there is no reason why a 1 percent chance that a witness is lying should trigger any less anxiety than a 1 (or .0000001) percent chance that someone else besides the defendant also shares the same fingerprint or alleles as the perpetrator. Although Heller makes a strong case that jurors greatly underestimate the probability that eyewitnesses or confessions are mistaken or lying, and overestimate the risk of error posed by “circumstantial” evidence, he doesn’t fully explain how these decision makers get beyond the “possibility” of witness inaccuracy to a *certainty* that the truth is known. Heller notes that “the certainty effect says that jurors ‘overweight outcomes that are *considered* certain,’ not outcomes that *are* certain.”²³² But Heller does not predict when a false certainty will or will not arise, especially in the face of steps—cross-examination, closing argument and the

²²⁶ Id. at 268; see supra notes 198-199 and accompanying text.

²²⁷ Heller, supra note 18, at 268.

²²⁸ Id. at 282 (“Jurors believe that they ‘should make accurate determinations with respect to the actual guilt or innocence of the defendant,’ . . . experience considerable stress in trying to comply with that self-imposed mandate [and] consistently report that choosing a verdict is the most stressful aspect of a criminal case.” (citations omitted)).

²²⁹ Id. at 283 (citations omitted).

²³⁰ Id. at 284-85.

²³¹ See supra notes 218-224 and accompanying text.

²³² Heller, supra note 18, at 285 (emphasis added).

heterodox perceptions of a dozen demographically diverse jurors²³³—designed to rub jurors' noses in fact of uncertainty.

Part of the problem, we believe, lies with Heller's distinction between "direct" evidence, mainly eyewitness testimony and confessions, and "circumstantial" evidence, including fingerprints and DNA. The more telling distinction, we believe, is between what we have called "big" evidence, including DNA and fingerprints as well as eyewitness testimony and confessions, and "small" evidence such as non-exclusionary non-matches. More precisely, the difference is between evidence with opposite profiles in regard to the Bayesian likelihood ratio. On the one hand is evidence that leads jurors to anchor on a high numerator probability that the evidence is present because the defendant is guilty, and to ignore the denominator possibility that the evidence appears though the defendant is innocent. On the other hand is evidence that is so likely to be present under all circumstances that the high denominator value obscures the fact that the numerator probability is even higher. At the limit, the difference is between two kinds of evidence that (if either actually existed) would provide lay decision makers with the irresistible security of true *certainty*. The first is evidence that lay decision makers assess as having a high number in the numerator and a zero in the denominator. This condition would characterize a "unique" and certain trait of the perpetrator that the defendant shares. The second is evidence lay decision makers assess as having a high denominator and a zero in the numerator. This condition characterizes traces that are likely to arise in the regular course of everyday life and have no relation to the crime—a dust mote found at an outdoor crime scene. In the next section, we hypothesize that the certainty intuitive decision makers crave leads them to embrace what others have exposed as the powerful myth of "uniqueness" or "individualization" in regard to "big" evidence," placing it in the former of these two imaginary categories, with a high numerator and no denominator. In the following section, we hypothesize the opposite craving, namely,

²³³ See, e.g., Shari Seidman Diamond, Truth, Justice, and the Jury, 26 Harv. J.L. & Pub. Pol'y 143, 153 (2003) ("Juries that begin deliberations holding heterogeneous verdict preferences tend to have more in-depth deliberations than juries that begin with a more homogeneous view of the evidence.").

the comforting ability to treat individually “small” evidence as entirely irrelevant, with a high denominator and no numerator.

4. *The “Uniqueness Fallacy.”* — The “pseudocertainty” effect reveals that intuitive decision makers can be persuaded that certainty exists when it doesn’t.²³⁴ We hypothesize that intuitive decision making might itself do the persuading. Note that the certainty effect is not so much a cognitive bias that generates demonstrable mistakes as an irrational preference that leads to sub-optimal outcomes. The dynamic is the one Samuel Johnson famously associated with second marriage, the triumph of hope over experience. Johnson referred not to an unrealized mistake but to an advertent preference for what one would like to be true over an opposite possibility that one knows is probably true.²³⁵

If the “certainty” achieved by reducing risk to zero creates vastly more advertent pleasure for people than other, comparable or greater, reductions of risk,²³⁶ then it would not be surprising if our brains, preferring hope over experience, look for ways to obtain the security inherent in zero-risk situations by finding certainty where it doesn’t exist. This could explain why individuals respond favorably to advertised pseudocertainty, when risks obviously remain. When identity is the issue, a particularly powerful way to achieve certainty is to conclude that characteristics matching perpetrator P and suspect S are unique, and thus that S *is* P. There thus may be important psychological benefits to believing in the truth of that equation, triggering the “certainty” bounce, even in the face of a patently more accurate judgment that generates an anxious indecision, namely, that “there is a strong probability that S is P, but a real possibility that he is not.” We know jurors agonize over the making a mistake that leaves a killer at large or convicts an innocent person, giving them every incentive to find shelter in even a false sense of certainty.²³⁷

The Bayesian equation (prior odds x likelihood ratio = subsequent odds) again helps model the heuristic process we describe. The values in the likelihood ratio (the probability that evidence would exist

²³⁴ See supra note 223 and accompanying text.

²³⁵ J. Boswell, *The Life of Samuel Johnson*, LL.D. 397 (1874, originally published 1791).

²³⁶ See supra note 218-224 and accompanying text.

²³⁷ See supra note 228 and accompanying text.

if the fact of consequence *were* true divided by the probability that the evidence would exist if the fact of consequence were *not* true) generate a measure of the probative weight of evidence through another equation:

$$(\text{numerator} - \text{denominator}) / \text{numerator} = \text{probative value.}^{238}$$

Notice that a high *numerator* is *necessary* to very high probative value but not sufficient for it. Very weighty evidence requires a low denominator as well. This construct suggests a shortcut the mind may take when it discerns uniqueness where none exists. A very high numerator—a high probability that evidence would exist if the defendant were, say, dangerous or guilty—triggers decisive action irrespective of the denominator, as if the denominator is zero or close to it. The shortcut is to “jump to the numerator conclusion” when the numerator is high and ignore or underestimate the denominator value. The result, we argue, is the uniqueness fallacy—to act decisively upon realizing that the defendant is behaving the way a guilty or dangerous person behaves without stopping to consider whether an innocent or benign person might also behave that way.

This dynamic helps explain the irrational impulse towards certainty triggered by “direct” evidence. Eyewitness testimony and confessions trigger the representativeness, simulation and confirmation biases, which in turn trigger a confidently high numerator probability that the evidence would exist if the defendant was guilty, which in turn triggers the cognitive economization of not wasting time considering the denominator. The result is to perceive the evidence as unique to the bad guy: a high numerator over a nonexistent denominator. Unlike Heller, however, we expect this phenomenon to accompany *all* high-numerator evidence, including DNA and fingerprints, as well as identifications and confessions.

One piece of evidence that human decision makers in fact anchor on large numerator probabilities to the exclusion of the denominator is Kahneman’s famous “Linda” experiment. Test subjects read a profile of “Linda,” a “thirty-one year[] old, single, outspoken, and very bright” woman, who was a

²³⁸ See Lempert, *supra* note 35, at 1047.

philosophy major in college, worries about “discrimination and social justice” and “participated in antinuclear demonstrations.” Subjects were then asked to pick the most likely future for Linda off of a list. Two futures listed were that “Linda is a bank teller who is active in the feminist movement” and that “Linda is bank teller.” Most subjects committed the logical fallacy of believing that the former probability is greater than, though the category of “bank tellers” is perforce larger because it contains, the category of “bank tellers active in the feminist movement.”²³⁹ Using our Bayesian model, it appears that the proper answer to this problem requires observers to add the numerator (the evidently large probability that a bank teller *concerned with social justice* is a feminist) to the denominator (the lower probability that the *average* bank teller is a feminist), but that most anchor on the former. Even as accomplished a scientist as Stephen Jay Gould admits falling prey to this fallacy: “I know [the right answer], yet a little homunculus in my head continues to jump up and down, shouting at me—‘but she can’t just be a bank teller; read the description!’”²⁴⁰ We hypothesize that Gould’s homunculus is the uniqueness fallacy—the intuitive tendency of the very high numerator to control, though it should be logically obvious that adding even a tiny denominator will produce a probability higher than the numerator by itself.

Also suggesting that decision makers pay closer attention to the numerator than the denominator probability is a study by Beyth-Marom and Fischhoff.²⁴¹ They asked subjects to identify information they would like to have when assessing the likelihood that a man drawn at random from a list of business executives and professors was a professor given that the man was a member of the Bear’s Club. Among the choices of information subjects might ask to use was the percentage of professors who are members of the Bears Club (in Bayesian terms, the numerator probability) and the percentage of business executives who are members of the Bears Club (the denominator probability). Most subjects wanted to know the former probability; many fewer cared about the latter one.²⁴²

²³⁹ Daniel Kahneman & Amos Tversky, Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment, 90 *Psych. Rev.* 293, 297-300 (1983).

²⁴⁰ Stephen Jay Gould, Bully for Brontosaurus: Further Reflections on Natural History 46 (1992).

²⁴¹ R. Beyth-Marom & Baruch Fischhoff, Diagnosticity and Pseudodiagnosticity, 45 *J. Personality & Soc. Psych.* 1185 (1983).

²⁴² *Id.* at 1193.

The “uniqueness fallacy” also appears to be a concern in everyday trials. Consider Federal Rules of Evidence 404-411. The first of these rules forbids jurors to rely on an inference all of us draw every day: the propensity, or representativeness, inference of action in conformity with a trait of character inferred from prior bad acts.²⁴³ The rest of these rules forbid jurors to rely on an inference of guilt or liability from evidence of a consciousness of guilt on the part of some actor—for example, someone who follows-up an accident with a so-called “subsequent remedial measure,”²⁴⁴ offers to settle a civil claim,²⁴⁵ pays the medical expenses of an accident victim,²⁴⁶ cops a plea to a crime,²⁴⁷ or insures himself against liability for accidents.²⁴⁸ In the same category are evidence doctrines discouraging inferences of a consciousness of guilt from silence in the face of a criminal accusation²⁴⁹ or from a suspect’s flight from arresting police officers²⁵⁰ or refusal to take a polygraph test.²⁵¹

Each of these rules is typically justified as a way to neutralize jurors’ tendency to jump to the conclusion that someone who did something bad in the past is likely to offend again or that people who are guilty act guilty, without considering innocent explanations for the behavior.²⁵² Rephrased in Bayesian terms, the law assumes jurors will treat the evidence as a confession of guilt and (via the representativeness, simulation and other biases) erroneously jump to a conclusion based on the high numerator without considering a not-inconsequential denominator probability. The law excludes the evidence, fearing that otherwise jurors will treat it as *unique* to guilty people—as having a high numerator value and a denominator worth no attention. Because blameworthy people so often take remedial measures, cover their tracks, run away or stay silent, the law expects jurors to assume that anyone who

²⁴³ Fed. R. Evid. 404(a)(1), 404(b)(1).

²⁴⁴ Fed. R. Evid. 407.

²⁴⁵ Fed. R. Evid. 408.

²⁴⁶ Fed. R. Evid. 409.

²⁴⁷ Fed. R. Evid. 410.

²⁴⁸ Fed. R. Evid. 411.

²⁴⁹ See, e.g., *Doyle v. Ohio*, 426 U.S. 610, 619 (1976) (barring inferences of guilt from silence following arrest and *Miranda* warnings); *Griffin v. California*, 380 U.S. 609, 615 (1965) (barring inferences of guilt from an accused’s failure to testify).

²⁵⁰ See e.g., Kenneth S. Broun et al., *McCormick on Evidence* 457-58 (6th ed. 2006).

²⁵¹ See, e.g., *deVries v. St. Paul Fire & Marine Ins. Co.*, 716 F.2d 939, 945 (1st Cir. 1983) (holding inadmissible witness’s refusal to take polygraph exam); *Aetna Insur. Co. v. Barnett Brothers, Inc.*, 289 F.2d 30, 34 (8th Cir. 1961) (same). See generally Broun et al., *supra* note 250, at 454-65 (noting evidence rules’ resistance to inferences of a consciousness of guilt from conduct).

²⁵² See, e.g., Lempert et al., *supra* note 81, at 280-81, 336-37 (discussing evidence rules designed to counteract jurors’ tendency to overestimate the value of evidence of conduct in which guilty people often engage).

has done one of these things is guilty and ignore the fact that innocent people do them, too—for example, the careful person who quickly repairs an unanticipated hazard revealed by a freak accident or the innocent person who worries that a polygraph will mistake nervousness for guilt.

Notice two things about the triggers for these common forms of juror misestimation: First, as Heller predicts, the simulation and (we would add) the representativeness biases are strongly at play.²⁵³ The forbidden evidence either reveals a trait resembling that of a guilty person (prior bad acts) or simulates the endgame of many crimes (actions to avoid apprehension). Second, contrary to what Heller predicts,²⁵⁴ every one of these examples is triggered by “circumstantial”—not “direct”—evidence: by inferences of a consciousness of guilt from action that has multiple interpretations. Clearly, the law assumes from long experience that human decision makers are disposed to turn what obviously is circumstantial or probabilistic evidence into unique evidence.

To be sure, as Heller argues, the same thing happens with eyewitness identifications and confessions, which are thought to be “direct,” “individual” and “unique” but aren’t. Contrary to Heller’s assumption, however, they are “circumstantial” not *only* because they depend upon probabilistic inferences of witness credibility.²⁵⁵ They are circumstantial, as well, because they depend upon aggregations of many only modestly “probable” individual matches.²⁵⁶ An eyewitness identification is powerful because the suspect matches multiple known attributes of the perpetrator—any one of which (e.g., small eyes or bushy eyebrows) is uninteresting. A confession is powerful because the details of the confessor’s story match so many of the known details of the crime, any one of which (a gun fired, someone screamed) is uninteresting.

The same thing also happens, however, when none of the heuristic biases Heller discusses applies and yet a disposition arises to treat merely probabilistic matches between traces associated with a crime

²⁵³ See supra notes 185-207 and accompanying text.

²⁵⁴ See supra note 227-229 and accompanying text.

²⁵⁵ See Heller, supra note 18, at 247 (assuming that credible eyewitness identifications and confessions have a probability of 1, or 100 percent).

²⁵⁶ See supra notes 77-91 and accompanying text.

and a suspect as if they involve a unique trait of a single human being. This occurs, for example, when DNA is treated as a “genetic fingerprint” though it is powerful only because of a non-unique aggregation of traces, each of which is no more telling than the fact that both the perpetrator and defendant are balding or left-handed.²⁵⁷ Recent scholarship also criticizes the myth of “uniqueness” and “individualization” as to tool marks, handwriting, bite marks, shoe prints and fingerprints.²⁵⁸ Fingerprints are especially interesting. Although powerful only because of a confluence of many individually uninteresting matches of lines and intersections, fingerprints are so reflexively thought of as “unique” that “fingerprint” and “unique” are dictionary synonyms.²⁵⁹ Our “uniqueness fallacy” thus explains what the biases Heller describes cannot fully elucidate: why criminal process decision makers overvalue not only narrative or “direct” evidence that invokes the simulation bias, but also entirely circumstantial evidence, such as drops of blood and mazes of lines left by oily human hands, that do not invoke that bias. The “uniqueness fallacy” fools jurors into treating all such “big” evidence as if it captured a “unique,” certainty-assuring property of the perpetrator, when it does not.

5. *The “Irrelevance Fallacy.”* — The flip side of the “uniqueness fallacy” we hypothesize, which disposes naïve decision makers to perceive uniqueness where none exists, is an “irrelevance fallacy” that leads them to assume that the weight of “small” evidence—non-exclusionary non-matches, for example—is so small that the evidence bears no consideration at all. There are hints of this fallacy in a study of the

²⁵⁷ See supra note 99-102 and accompanying text. See also Jonathan J. Koehler, When Are People Persuaded by the DNA Match Statistics?, 25 L. & Hum. Behav. 493, 508-09 (2001) (finding that when told that the probability of a coincidental match between DNA found at a crime scene and a suspect’s DNA is low, mock jurors assume the probability of a coincidence or error is zero); Levy, supra note 95, at 26 (noting misimpression of DNA as “a genetic fingerprint” that “is individually specific” and “does not belong to any other present or future person on earth”); Neufeld & Colman, supra note 76, at 50 (decrying misimpression that DNA “identifies the ‘genetic code’ unique to an individual” and “is as unique as a fingerprint”); Smith et al., supra note 25, at 410, 414 & nn.9-11, 15, 16 (noting that jurors’ tendency to “under-value probabilistic evidence when compared to a Bayesian calculation” does not hold for DNA).

²⁵⁸ See, e.g., Cole, supra note 84, at 234-35, 242 (juxtaposing a so-called “fingerprint examiner’s fallacy” to “broad” scholarly agreement “that individualization and uniqueness [do not] constitute the proper conceptual foundations for forensic identification” (citing sources)); Simon A. Cole et al., Beyond the Individuality of Fingerprints: A Measure of Simulated Computer Latent Print Source Attribution Accuracy, 7 L. Prob. & Risk 165 (2008); Michael J. Saks & Jonathan J. Koehler, Individualization Claims in Forensic Science. Still Unwarranted, 75 Brooklyn L. Rev. 1187, 1191-94 (2010); Michael J. Saks & Jonathan J. Koehler, The Individualization Fallacy in Forensic Science Evidence, 61 Vand. L. Rev. 199, 203-08 (2008) (criticizing “individualization fallacy”).

²⁵⁹ Compare Random House Webster’s College Dictionary, <http://www.definitions.net/definition/Fingerprint> (defining fingerprint as “any unique or distinctive pattern that presents unambiguous evidence of a specific person, substance, disease, etc.”) with supra note 78 (explaining why fingerprint evidence is not unique).

weight subjects gave to a match between the rare blood-type and hair characteristics of an unknown perpetrator and defendant. The authors were “most surprise[ed]” by “how easily people can be persuaded to give no weight” at all to such evidence when presented with the so-called “Defense Attorney’s Fallacy,” that “even though only two percent of the population has characteristic X, in the entire population of this city, there are hundreds of people with that rare trait.”²⁶⁰ This is a fallacy because the realistic number of suspects in most cases is smaller than the entire population, leaving it unlikely that any member of that smaller group has the rare trait.

Consider, as well, how the numerator-focused dynamic we describe above likely works against defendants relying on “small” non-matches. Assume that police find a partially smoked cigarette at the scene of a crime with which D is charged, and that D is a non-smoker. In considering the likelihood ratio associated with the cigarette butt—a ratio D claims has a larger denominator than numerator, meaning the evidence tends to prove *innocence*—the jury will quickly see that the numerator is quite large. The probability of finding a cigarette butt at a crime scene is high whether or not the perpetrator smokes. Given the high numerator, the jury may fail to realize that the denominator is modestly larger, because it includes all of the innocent ways the butt could have gotten there, plus one *guilty* way: being left behind by a smoker-culprit who is *not* the defendant. Indeed, a disposition to ignore the denominator here is very nearly the bias revealed by the “Linda” study.²⁶¹ The evidence is weak, to be sure, because the difference between the numerator and denominator is small. But the irrelevance fallacy suggests the jury will conclude that the evidence would be present even if the non-smoker defendant was guilty (anchoring on the numerator) and ignore—treat as irrelevant—the slightly higher denominator. If D tries to trigger the simulation effect by imagining for the jury a scenario in which the affray dislodged the cigarette from the mouth of a smoker-perpetrator who is not the defendant, the prosecutor can easily counter with far more scenarios in which a non-perpetrator dropped the butt.

²⁶⁰ Thompson & Schumann, *supra* note 144, at 182.

²⁶¹ See *supra* notes 239-240 and accompanying text.

Now, assume there are two such non-matches—not only the cigarette butt but a stray maroon button at the scene that is not associated with the victim’s clothing or the defendant’s when he was arrested right after the crime. As we develop above, although both non-matches are weak evidence of innocence, together, they can gain probative steam—if the jury will consider them together, aggregating the individually small probability of innocence associated with each. If, however, the irrelevance fallacy leads jurors to reject each non-match as unworthy of their cognitive time, there are only zeros to aggregate.

Of interest here are recent studies by Lisa L. Smith and colleagues, who asked jury-eligible subjects to rank the value, as proof of guilt, of certain DNA, fingerprint and bloody-shoeprint evidence found at or near a crime scene that matched a suspect.²⁶² Each subject considered twelve examples that differed in terms of whether the evidence was highly “relevant” (likely to have been left at the scene by the perpetrator) or weakly “relevant” (many innocent explanations for the suspect’s blood, fingerprint or footprint being there), and whether the evidence was highly “mobile” (easily could have been migrated to the crime scene apart from the crime) or not very “mobile” (evidently was left during the crime). Based on these features, the authors created four categories of evidence (high relevance/low mobility, low relevance/low mobility, etc.) that ranged from being likely to reveal a match of a “unique” trait to being likely to be coincidental or “irrelevant.”²⁶³ Evincing considerable facility with circumstantial evidence, subjects drew sharp and appropriate distinctions between the evidence, recognizing that some was strong, some was weak and some was in between.²⁶⁴

In a second study, the researchers presented three groups of subjects with descriptions of evidence at a murder trial. Each group received the same evidence (mainly witness statements) with one exception. One group was told that strong DNA evidence (high relevance/low mobility) linking the defendant to the crime was also present. Another group was told that moderately probative DNA evidence

²⁶² Smith et al., *supra* note 25.

²⁶³ *Id.* at 410-11.

²⁶⁴ *Id.* at 411.

(high relevance/high mobility) was present. A third group was told that weak DNA evidence (low relevance/high mobility) was present. The mock jurors estimated the strength of the evidence after hearing all the evidence, and again after hearing closing statements. At the latter stage, the subjects also rendered a verdict.²⁶⁵ Again, the three groups of mock jurors appropriately distinguished the cases based on the relative strength of the forensic evidence—estimating a slightly higher probability of guilt in the *strong* than in the *moderately strong* DNA cases, and a much smaller probability of guilty in the *weak* DNA case.²⁶⁶ So far the news is all of the dog-bites-man sort.²⁶⁷

The study ends, however, on a man-bites-dog note. In rendering verdicts, the mock jurors were much more likely to find the strong-DNA defendant guilty (40%) than the moderately strong DNA defendant (23%), though they rated the probability of guilt about the same in the two cases. And they were not that much more likely to find the defendant guilty in the moderately strong DNA case than in the weak DNA case (15%), though they recognized that the probability of guilt was much greater in the former case.²⁶⁸ Notwithstanding the mock jurors' rational estimates of the probability of guilt based on differences in the strength of the DNA evidence, when reaching a verdict, the subjects paid less attention to the probabilities than to the ease with which they could reach a "uniqueness" conclusion (proving guilt) or could identify alternative explanations for the DNA evidence and ignore it (leading to acquittal).²⁶⁹ This may provide evidence of the uniqueness fallacy in the former case and the irrelevance fallacy in the latter one.

An example of how the uniqueness and irrelevance fallacies can lead law enforcement astray in a case involving "big" circumstantial evidence of guilt and "small" non-match evidence of innocence is the

²⁶⁵ Id. at 411-12.

²⁶⁶ Id. at 412-13.

²⁶⁷ See supra notes 227-229 and accompanying text.

²⁶⁸ Smith et al., supra note 25 at 412-13.

²⁶⁹ See also Sonnemans & van Dijk, supra note 209, at 27-28 (reporting studies showing that jurors are good at estimating relative probabilities but tend to ignore them when reaching a verdict, jumping to stronger conclusions than warranted). The uniqueness and irrelevance fallacies help explain why jurors overvalue "unique"-seeming eyewitness identifications, see supra notes 225-226 and accompanying text, but undervalue eyewitnesses' inability to identify the defendant, see Hunter A. McAllister & Norman J. Bregman, Juror Underutilization of Eyewitness Nonidentifications: Theoretical and Practical Implications, 71 J. Applied Psych. 168, 169-70 (1986).

FBI's material-witness arrest of Brandon Mayfield. Mayfield was an Army veteran and family lawyer in suburban Portland, Oregon, who had married an Egyptian woman and converted to Islam.²⁷⁰ With great fanfare, the FBI arrested Mayfield and held him for nineteen days as a material witness in connection with the March 2004 Al Qaeda bombings of the Madrid commuter rail system.²⁷¹

The FBI based Mayfield's arrest on a "100% verified" match between his fingerprint and partial prints found by Spanish authorities on a bag of detonating devices shortly after the explosions.²⁷² The match initially overwhelmed several non-exclusionary non-matches: Mayfield had no known expertise as a bomb-maker or access to bomb-making equipment or terror suspects, had not left the country since 1994, and was rejected as a suspect by Spanish authorities who contested the fingerprint match and identified an Algerian immigrant with no connections to the United States as the likely bomber. Only after Spanish officials examined and rejected Mayfield's fingerprints a second time did the FBI release him, though it continued to subject him and his family to intensive surveillance.²⁷³

Mayfield ensuing lawsuit unearthed FBI records revealing that it had identified twenty Americans whose fingerprints were "similar" to the Madrid prints, triggering surveillance of all twenty. Mayfield's name quickly topped the list,²⁷⁴ probably because he was married to an Egyptian woman, converted to Islam, provided modest legal services to a member of a group of people attempting to travel to Afghanistan to work with the Taliban, and supported various do-gooder causes.²⁷⁵ The FBI subsequently

²⁷⁰ See Stephen T. Wax, *Kafka Comes to America: Fighting for Justice in the War on Terror* 6 (2008).

²⁷¹ See *Mayfield v. United States*, 504 F. Supp. 2d 1023, 1026-27 (D. Or. 2007) ("On March 11, 2004, in Madrid, Spain, terrorists' bombs exploded on commuter trains, murdering 191 persons, and injuring another 1600"); Susan Schmidt, *American Held in Madrid Bombings*, Wash. Post, May 7, 2004, at A1.

²⁷² "Using standard protocols and methodologies, FBI fingerprint examiners determined that the latent fingerprint [from Spanish authorities] was of value for identification purposes" and was "linked to Brandon Mayfield," a result that was "independently . . . confirmed by an outside experienced fingerprint expert." May 24, 2004 Statement of the FBI, <http://www.fbi.gov/news/pressrel/press-releases/statement-on-brandon-mayfield-case>. See Steven T. Wax & Christopher J. Schatz, *A Multitude of Errors: The Brandon Mayfield Case*, *The Champion*, Sept. /Oct. 2004, at 6 ("FBI Senior Fingerprint Examiner Terry Green manually compared the [twenty] potential matches with the digital image of Latent Fingerprint No. 17 and found a '100 percent' match with [one] print on the list [belonging to] Brandon Mayfield.").

²⁷³ U.S. Dep't of Just., Office of the Insp. Gen., *A Review of the FBI's Handling of the Brandon Mayfield Case 230* (March 2006), <http://www.justice.gov/oig/special/s0601/Chapter1.pdf>.

²⁷⁴ *Mayfield v. United States*, 504 F. Supp. 2d at 1027.

²⁷⁵ See Wax, *supra* note 270, at 6-7.

settled the lawsuit for \$2 million, apologized to Mayfield and blamed the mishap on “an unusual similarity” between Mayfield’s fingerprint and a copy of a print associated with the Madrid bombings.²⁷⁶

Mayfield’s case was rife with heuristic traps. The representativeness and simulation biases associated with Mayfield’s conversion to Islam and left-of-center sympathies evidently triggered a powerful, if imaginary, scenario of a turn towards Islamic terrorism that blinded officials to the base rate of nineteen other suspects in the FBI database with “similar” fingerprints, not to mention scores of additional suspects that a simple extrapolation of that number to international databases would have suggested. Via the confirmation bias, the “100%”-certain match and accompanying scenario in turn invoked the uniqueness fallacy that the fingerprint was personal to Mayfield and the irrelevance fallacy as to the various, non-exclusionary non-matches. Apart from the fortuity of parallel Spanish and U.S. investigations, it is unclear whether or how quickly Mayfield would have been cleared of the many capital crimes of which he was suspected.

Supplementing the many studies of cognitive fallacies is a growing literature on practical steps institutions, the justice system included, do or could use to head off debilitating biases.²⁷⁷ Above, we note how Evidence Rules 404-411 diminish the representativeness bias and tendency to overvalue the numerator and undervalue the denominator in assessing evidence.²⁷⁸ Likewise, “devil’s advocate”

²⁷⁶ Statement of the FBI (Jan. 6, 2006), <http://www.fbi.gov/news/pressrel/press-releases/fbi-responds-to-the-office-of-inspector-general2019s-report-on-the-fingerprint-misidentification-of-brandon-mayfield>. See *Mayfield v. United States*, 599 F.3d 964, 968 (9th Cir.), cert. denied, 131 S. Ct. 503 (2010) (describing FBI’s agreement to pay compensatory damages of \$2 million, destroy fruits of its surveillance and apologize); Apology Note, Washington Post, Nov. 29, 2006, http://www.washingtonpost.com/wp-dyn/content/article/2006/11/29/AR2006112901155.html?nav=rss_nation/nationalsecurity.

²⁷⁷ See, e.g., Burke, *Prosecutorial Decision Making*, supra note 208, at 1613-31 (proposing procedures prosecutors and courts can use to mitigate confirmation bias in deciding whom to charge and whether to produce exculpatory evidence); Colin Camerer & Ari Vepsäläinen, *The Economic Efficiency of Corporate Culture*, 9 *Strategic Manag. J.* 115 (1988); Lynne L. Dallas, *Does Corporate Law Protect the Interests of Shareholders and Other Stakeholders?: The New Managerialism and Diversity on Corporate Boards of Directors*, 76 *Tul. L. Rev.* 1363, 1407 (2002) (discussed infra note 281); Hammond et al., *Improving Scientists’ Judgments of Risk*, 4 *Risk Analysis* 69, 72-77 (1984); Christine Jolls & Cass R. Sunstein, *Debiasing through Law*, 35 *J. Legal Stud.* 199, 212 (2006) (arguing for legal policies that help eliminate problems generated by “bounded rationality” and cognitive “biases”); Kahneman, *Thinking, Fast*, supra note 184, at 417-18 (“Organizations are better than individuals when it comes to avoiding [heuristic] errors, because they naturally think more slowly[,] . . . have the power to impose orderly procedures [and] can institute and enforce the application of checklists, as well as more elaborate exercise, such as reference-class forecasting and postmortem.”); Jesse M. Pines, *Profiles in Patient Safety: Confirmation Bias in Emergency Medicine*, 13 *Academic Emerg. Med.* 90 (2006) (listing ways hospital emergency physicians can combat confirmation bias); Michael J. Saks & Robert F. Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 *L. & Soc’y Rev.* 123, 131 (1980); Thaler & Sunstein, supra note 222, at 177-78.

²⁷⁸ Sees supra notes 243-248 and accompanying text.

mechanisms (long used by the Catholic Church to increase the reliability of canonization rituals²⁷⁹) can mitigate simulation, certainty and confirmation bias. Such mechanisms force people to imagine scenarios and present arguments that counter their initial dispositions,²⁸⁰ or assign decisions to large and diverse groups likely to generate competing scenarios on their own (a strategy the Constitution arguably adopts by requiring juries of at least six people chosen from venirees representing the entire community²⁸¹). This literature raises two questions: Do our existing legal procedures and rules sufficiently counteract the biases and fallacies we have discussed? If not, are there other measures that could succeed? The next two sections conclude that, on balance, existing procedures and rules aggravate the problem. Part VI suggests measures that might have an ameliorative effect.

B. Structural Disadvantages

1. *Reasons to Doubt the Adversarial Antidote.* — By constantly confronting jurors with counter-scenarios and competing arguments, adversarial processes are supposed to provide an effective antidote to the heuristic biases described here.²⁸² To whatever extent investigators, forensic analysts and prosecutors may commit themselves, including via heuristic biases, to the “whodunit” theory the state offers at trial, the defense commits itself, and works to commit jurors, to equally vivid alternative possibilities. Sorting

²⁷⁹ *Advocatus Diaboli*, in 1 *The Catholic Encyclopedia* (1907). See, e.g., Craig A. Anderson & James J. Lindsay, *The Development, Perseverance, and Change of Naïve Theories*, 16 *Soc. Cognition* 8, 24 (1998) (using “counter-explanation” process, in which the subject “imagines and explains how a different [scenario] is (or might be) true” to counteract confirmation bias); Burke, *Prosecutorial Decision Making*, supra note 208, at 1618 (encouraging prosecutors to avoid cognitive bias “by generating pro-defense counterarguments to [their] own . . . interpretations of the evidence”); Michael R.P. Dougherty et al., *The Role of Mental Simulation in Judgments of Likelihood*, 70 *Org. Behav. & Hum. Dec’n Processes* 135, 136 (1997); Lord et al., supra note 208, at 1238-40 (causing experimental subjects to consider the opposite of their initial hypothesis induced more accurate results than instructing them to be fair and unbiased).

²⁸⁰ See, e.g., Findley & Scott, supra note 208, at 371; Adam D. Galinsky & Gordon B. Moskowitz, *Counterfactuals as Behavioral Primes: Priming the Simulation Heuristic and Consideration of Alternatives*, 36 *J. Exper. Soc. Psych.* 384, 391 (2000) (noting the ameliorative effect of “priming” study subjects with plausible alternatives); Keith E. Niedermeier et al., *Jurors’ Use of Naked Statistical Evidence: Exploring Bases and Implications of the Wells Effect*, 76 *J. Personality & Soc. Psychol.* 533, 541 (1999) (similar).

²⁸¹ *Ballew v. Georgia*, 435 U.S. 223, 231 n.10 (1978) (concluding, based on studies showing that the chance of convicting an innocent person increases as the number of jurors decreases, that criminal juries of less than six persons violate due process (citing studies)); supra note 233. See also Dallas, supra note 277, at 1402 (arguing that heterogeneous corporate boards are less susceptible to confirmation bias because they harbor more competing views).

²⁸² See, e.g., Lon L. Fuller, *The Adversary System*, in *Talks on American Law* 30, 39-40 (H. Berman ed. 1971) (“An adversary presentation seems the only effective means for combating the natural human tendency to judge too swiftly in terms of the familiar that which is not yet fully known.”); John Thibaut, et. al., *Adversary Presentation and Bias in Legal Decisionmaking*, 86 *Harv. L. Rev.* 386, 397 (1972) (arguing that adversarial procedures counteract decision maker bias more effectively than inquisitorial procedures).

through the competing presentations is expected to induce jurors to focus on relevant aspects of the base rate, numerator and denominator in resolving the dispute. There are two reasons, however, why adversarial procedures do not provide effective antidotes to the biases against aggregative analysis of non-unique evidence.

First, even experts trained to recognize the ill-effects of cognitive biases unwittingly succumb to them.²⁸³ We should not assume, therefore, that lawyers are immune and can effectively wean jurors from these errors. Even if both sides are equally prone to mistakes, there is no reason to expect the mistakes as a whole to neutralize each other in regard to the search for the truth in a given case.

Second, the opposing sides of a criminal case are unlikely to be similarly situated in relation to the representativeness, simulation, confirmation, certainty, uniqueness and irrelevance heuristic advantages that we describe. A party blessed with eyewitness testimony, a fingerprint or some other kind of “big” evidence that obscures the aggregation of many only modestly probative matches and triggers the uniqueness fallacy has numerous heuristic advantages. Even if the other side has an equally strong case from a Bayesian perspective, but has to make it by aggregating the force of many bits of “small” evidence that call forth no similar sense of uniqueness and invite a series of irrelevance fallacies, that party is disadvantaged.²⁸⁴ Indeed, as the next sections demonstrate, together with the various heuristic fallacies, the state’s monopoly over crime scene evidence and first crack at key witnesses causes exactly this uneven distribution of advantage and disadvantage to recur, systematically favoring the prosecution over the defense.

2. *The Biasing Effect of the State’s Monopoly over the Initial Investigation.* — The Carlos DeLuna case²⁸⁵ illustrates the government’s monopoly control over initial criminal investigations and its uneven effect on the fight for heuristic advantage between prosecution and defense. The murder victim in the case, a store clerk, called 911 to report a “Mexican” man with a knife in her store. She was heard

²⁸³ See Kahneman, Thinking, Fast, supra note 184, at 415 (“[M]y intuitive thinking is just as prone to overconfidence, extreme predictions, and the planning fallacy as it was before I made a study of these issues.”); supra note 188 and accompanying text.

²⁸⁴ See supra notes 234-276 and accompanying text.

²⁸⁵ See supra notes 8-16 and accompanying text.

begging someone to take the store's money, screaming and struggling. When police arrived moments later, they found her near death from a knife wound through her lung that hemorrhaged blood as she resisted the killer's efforts to wrestle her into a back room.²⁸⁶

A police audiotape captured the 911 call and radio traffic during an ensuing manhunt for a man seen fleeing the store on foot. Prosecutors played the 911 tape at DeLuna's trial but misinformed defense counsel that the manhunt portion of the tape had been recorded over without being saved. It came to light only after DeLuna was executed, revealing that police had chased another man along a different path for thirty minutes before a call about DeLuna diverted attention to him. Multiple aspects of the dress, grooming and direction of flight of the man police initially chased matched a description they had from the only eyewitness to the crime of a mustachioed, shabbily dressed "derelict" who raced north and west behind the store. The description did not match the clean-shaven, well-dressed DeLuna who was seen two blocks east of the store seconds after the killing. At trial DeLuna testified that he had been with a man named Carlos Hernandez earlier that evening and had seen him attack the convenience store clerk. The prosecutor branded DeLuna a liar and Hernandez a "phantom," and the jury heard nothing more about Hernandez nor anything about his history of armed convenience store robberies and knife violence, physical resemblance to DeLuna and repeated admissions to stabbing the convenience store clerk. Nor did the jury learn that at least seven features of the eyewitness's description of the attacker that did not match DeLuna did match Hernandez.²⁸⁷

When police arrested DeLuna forty minutes after the crime, they decided against a station-house line-up and took him to the crime scene for nighttime show-up identification in a poorly lit parking lot. On the ride over, DeLuna told the arresting officer he didn't commit the crime but knew who did. Police never followed up or questioned DeLuna.²⁸⁸ When the squad car arrived at the convenience store, the eyewitness was too scared to view the suspect, but relented when officers surrounded him and let him

²⁸⁶ See Liebman et al., *supra* note 8, at ___.

²⁸⁷ See *id.* at ___; *supra* notes 10-16, 26 and accompanying text.

²⁸⁸ See Liebman et al., *supra* note 8, at ___.

view the seated, shirtless and hand-cuffed suspect through squad car window while flashlights were trained on his face. After observing DeLuna for fifteen seconds, the Anglo witness identified him as the culprit. Years later, the witness admitted he had trouble telling one Latino from another and said he was only “70 percent sure” of the identification, and would have been only “50 percent sure” if police hadn’t told him beforehand that they had found DeLuna cowering under a pick-up truck.²⁸⁹

Meanwhile, a lone detective and a photographer had entered the store where the killing took place. Evidence the detective overlooked, never tested forensically or failed to disclose to the defense is listed above.²⁹⁰ The detective found the murder weapon, wet with blood and flesh, but tested it for fingerprints at the scene with graphite powder, an improper technique for recovering prints from wet and oily surfaces. No prints were found, and the graphite prevented later lab testing with proper materials. None of the prints found elsewhere at the scene matched DeLuna. Shortly after learning that an eyewitness had identified DeLuna and barely an hour after the scene investigation began, the detective ended it and turned the store back to employees. They scrubbed it down overnight—wiping away a bloody handprint, unnoticed by police, on the inside of a window—and opened on time the next morning. No one ever saw the intact crime scene in daylight.²⁹¹

These facts suggest how heuristic biases may affect police officers’, forensic analysts’ and prosecutors’ exercise of their monopoly over the initial handling of vast amounts of potential evidence, and the heuristic advantage the monopoly affords the state. As we note above, law enforcement officials are themselves subject to the representativeness, simulation and other biases and instinctively expect jurors to be as well.²⁹² Because a case premised on unique-seeming “big” evidence linking the defendant to the crime appears to be far stronger than one relying on an aggregation of “small” evidence, a central law enforcement objective from the start is to use its exclusive first crack at the crime scene and witnesses to obtain “unique” evidence—a confession, eyewitness identification or DNA or fingerprint match. The

²⁸⁹ See *id.* at ____.

²⁹⁰ See *supra* text accompanying note 26.

²⁹¹ See Liebman et al., *supra* note 8, at ____.

²⁹² See *supra* note 187 and accompanying text.

DeLuna detective's blundering haste to find fingerprints on the knife is one example. Another is the lengths police went to secure an eyewitness identification: passing up a more reliable line-up in favor of a suggestive nighttime show-up, cluing in the witness to the compromising circumstances of DeLuna's arrest, and allowing the frightened witness to view the seated, shirtless and handcuffed suspect through the window of a squad car, while a circle of police officers provided protection against the presumed-guilty suspect while spotlighting his face with flashlights.²⁹³

If "big" evidence implicating a suspect does not immediately appear, police can keep looking for it without adversarial interference, which only formal charges can trigger. Once a "big"-evidence anchor for the state's case appears, the uniqueness fallacy diminishes investigators' disposition to look for additional "small" evidence matching the suspect, including because such evidence might *not* match him and weaken the state's case. Investigation that does take place tends, at least subconsciously, to "confirm" what the "big" evidence has simulated and "represents": the suspect's guilt.²⁹⁴ Meanwhile, the irrelevance fallacy renders invisible any non-matching evidence that already has accumulated. Given obvious reasons why the non-match could occur, though the defendant is guilty—given, that is, the high numerator value—the modestly greater denominator probability of innocence will be ignored, leaving nothing to aggregate with other such evidence into a reasonable doubt.²⁹⁵

In DeLuna, for example, when the detective ended her scene investigation upon hearing of the "big" show-up identification, she left behind numerous undiscovered bits of evidence of identity that may have matched DeLuna—or may not have matched him and even yielded an *exclusionary* non-match. We know about these lost items only by the happenstance of a rare post-execution investigation. They include a bloody handprint washed away by a store employee who no one interviewed at the time and several items spotted in previously undisclosed police photographs:

²⁹³ See Liebman et al., *supra* note 8, at ____.

²⁹⁴ See *supra* notes 208-214 and accompanying text.

²⁹⁵ See *supra* notes 238-259 and accompanying text.

- *A clump of hair.* The eyewitness told police the assailant at one point had the female victim by the hair, but this was never confirmed. Nothing was made at trial of the absence of foreign hair in DeLuna's fingernails and on his clothing at the time of his arrest.
- *Bloody shoeprints.* The victim was barefoot, so bloody shoeprints found inside the store and on the sidewalk outside were not hers. Police made no casts, close-ups or measurements of the unnoticed shoeprints and never compared them to DeLuna's blood-free tennis shoes.
- *A wad of chewed gum.* The gum evidently was disgorged during the struggle because it landed on a blood-stained calendar that itself was knocked onto the floor during the melee.
- *Cement or cinderblock shards* on a carpet mat where the struggle occurred. Lab technicians analyzed DeLuna's shoes but were not asked to look for rock fragments and reported none.²⁹⁶

Also unexplored were several non-exclusionary non-matches that had appeared before the scene investigation ended: DeLuna's claim that he knew who committed the assault; numerous non-matching aspects of the eyewitness description, including the culprit's dress, grooming and direction of flight; a maroon button at the scene that didn't match DeLuna's or the victim's clothes; and the absence of even a scintilla of blood on DeLuna's hands, clothing and hair, despite the bloodbath officers found at the crime scene.²⁹⁷

Once a suspect is arrested and counsel is appointed, the process veers sharply from inquisitorial to adversarial. But the "fair fight" the adversarial system imagines at this point still may not materialize. The government's monopoly over the initial investigation already will have allowed it, alone, to shape the first and best shot anyone will get at the crime scene and witnesses, with the goal of obtaining seemingly "unique" evidence of a suspect's guilt and all the heuristic advantages such evidence affords. As in *DeLuna*, the state may have made vital decisions about how many professionals to assign to the scene investigation, how long and thoroughly to search, how best to lift fingerprints, and what evidence to

²⁹⁶ See Liebman, *supra* note 8, at __; *supra* text accompanying note 26.

²⁹⁷ See Liebman et al., *supra* note 8, at __; *supra* note text accompanying note 26.

preserve and test. Eyewitnesses also will have been the state's alone at a point when their memories were freshest and things police said and procedures used forever colored the witnesses' view of what happened. After the state obtained unique-seeming evidence of a suspect's guilt, it may have allowed substantial amounts of additional "small" and even "big" evidence to go by the boards, such as bloody hand and shoeprints and a chance to apprehend and question alternative suspects. Unique access to the crime scene and the ability at will to end the investigation thus will have maximized the state's access to big evidence and heuristically powerful scenarios and let it choose when its theory likely "covered"²⁹⁸ the highest possible proportion of found small evidence, deterring a search for more. In important respects, that is, the state's monopoly over the initial investigation will have given it the exclusive power to determine whether much of the evidence—including "big" evidence pointing *away* from the state's prime suspect, as well as a plethora of "small" evidence—even *exists*.

Conversely, the defense may never have a shot at the crime scene or witnesses police failed to identify, and will speak to many found witnesses only after police have shaped their view of the matter.²⁹⁹ The only scenario-generating evidence to which an innocent defendant is assured access is whatever heuristically disadvantageous *small* non-matches law enforcement happened to turn up before shutting down the investigation, then decided to disclose under discovery rules built to apply only to *big* evidence.³⁰⁰ Even apart from chronic resource disparities, therefore, the innocent defendant will find it difficult to reconstruct the clues the state missed or withheld and, with so little to go on, will likely end up with *no* scenario to offer and only the weakest of defenses, namely, reasonable doubt.³⁰¹

There are other heuristic skews, as well. The very job of a detective and prosecutor is to simulate a "bad guy's" guilt, based, if possible, on evidence that seems to be "unique" to the defendant. Just by describing what the defendant did in an arrest warrant or indictment, the detective and prosecutor

²⁹⁸ See *supra* notes 199.

²⁹⁹ See *supra* note 289 and accompanying text. State's witnesses are not obliged to speak to defense counsel, and are often advised as such by police officers who themselves decline to talk to the defense.

³⁰⁰ See *infra* notes 335-352 and accompanying text.

³⁰¹ See *supra* note 206 and accompanying text.

solemnly commit to the validity of their evil-defendant theory. Doing so both fuels and feeds off of the simulation, confirmation and certainty biases, as well as the uniqueness fallacy as to any “big” evidence on which they rely. Defense lawyers, by contrast, exist to offer “alternative hypotheses.” The only proposition defense counsel need embrace is reasonable doubt. Because defense lawyers trade in the pallid probabilities necessary to avoid conviction, there is less chance the zeal they exercise for their clients will be fortified by the heuristic biases that feed police officers’ and prosecutors’ zealous protection of “the People” and “the State.” There thus is heuristic truth to stereotypes of the hard-charging white knight of a prosecutor pursuing evil without a hint of doubt about the justness of her cause, and the cynical defense lawyer representing guilty defendants with his fingers crossed behind his back.

Because many decisions to prosecute and all trials call for an up-or-down decision on a single suspect, not a comparison of all possible suspects, they present decision makers with a single, prominent theory encapsulated in a prior arrest warrant or in the indictment and state’s opening and closing arguments and case in chief. The question is not “Whodunit?,” but “Did the *Defendant* do it?” From the start, the adversarial process “simulates” the guilt of one person. In contrast to the numerous possible answers to the state’s initial “Whodunit?” inquiry, including, “We don’t know and are still looking,” the “Did the defendant do it?” drama the adversarial system presents has only two possible answers: the defendant did it, or the culprit escaped justice. Of those, the former is far more likely to feel like a success than the latter.

For these reasons, and because the obligations of both the prosecution and defense are satisfied by a belief that the state’s whodunit? narrative is probably true, almost everything about prosecutions and trials conspires to imbue the decider with the representativeness, simulation and confirmation biases. Those biases in turn invest “big,” seemingly “unique” evidence with a false sense of certainty and demean individually “small” non-matches as irrelevant, whatever their aggregate force. Once infected by heuristic bias, that is, key aspects of the adversarial process conspire to negate the chief protection it claims to offer defendants through the “beyond a reasonable doubt” standard.

The neural hard-wiring that prompts us to take cognitive shortcuts is a final reason to doubt the power of the adversarial system as currently implemented to neutralize them. The system assumes that, by constantly modeling a battle between sides committed to alternative truths (assuming the defense qualifies as such, given the above analysis), it can deter jurors from taking shortcuts and focus them intuitively on counterweights to bias such as alternative scenarios, base rates and denominator values slightly higher than numerators. But as we have seen, much of the heuristic damage occurs during the initial investigation before the adversarial system kicks in. Worse, even true “experts,” and surely jurors without their training, succumb to the heuristic homunculus jumping up and down in even Stephen Jay Gould’s head.³⁰²

The question remains whether other existing legal doctrines, such as the evidentiary rules mentioned above,³⁰³ true the balance, knocking “big,” seemingly “unique” evidence down to size and adding clout to seemingly “irrelevant” non-exclusionary non-matches. In the next section, we identify several existing doctrines that actually aggravate the problem. But all is not lost. In the following Part, we identify steps to achieve the desired effect.

C. Legal Obstacles

Apart from evidence that is irretrievably lost or wiped away, one might expect legal rules to compensate for structural failings and blind spots. But far from curing the heuristic inflation of “big” evidence and deflation of aggregations of “small” evidence, legal rules affecting each stage of a criminal case exacerbate or succumb to these mistakes. We begin with a common legal categorization of evidence that epitomizes and institutionalizes the problem. Then, we identify additional offending rules at each stage of the criminal process.

1. *Rules Distinguishing “Direct” and “Circumstantial” Evidence.* — The law does not recognize a distinction between “big” and “small” evidence or privilege one over the other. But jury instructions in

³⁰² See supra notes 188, 240 and accompanying text.

³⁰³ See supra notes 243-248 and accompanying text.

nearly all jurisdictions draw a related distinction between “direct” and “circumstantial” evidence.³⁰⁴ Connecticut Jury Instruction 2.4-1 juxtaposes “two kinds of evidence, direct and circumstantial,” the former defined as “testimony by a witness about what that witness personally saw or heard or did,” and the latter being “indirect evidence from which you could find that another fact exists, even though it has not been proved directly.”³⁰⁵ California distinguishes “direct” evidence that “can prove a fact by itself” from “circumstantial evidence that does not directly prove the fact to be decided, but is evidence of another fact or group of facts from which you may conclude the truth of the fact in question.”³⁰⁶

Although following the modern trend, both instructions say there is “no legal distinction between direct and circumstantial evidence as far as probative value” and let jurors “give equal weight to both,”³⁰⁷ the instructions still tend to confuse jurors and convey the opposite sense. Both instructions juxtapose an encomium, “direct,” and the often pejorative “indirect,” and suggest that the former is more straightforward (revealing what actually happened or what the witness “personally saw,” compared to “evidence from which you *could* find that another fact exists, *even though* it has *not been proved directly*”) and more powerful (“prov[ing] a fact by itself” versus “not directly prov[ing] the fact”).³⁰⁸ No wonder, then, that jurors continue to associate “direct” with “strong” and “circumstantial” with “weak” evidence.³⁰⁹ In one study, subjects presented with scenarios involving strong circumstantial evidence (for example, phone records or expert comparisons of paint marks, which most subjects mistakenly defined as “direct”) or weak direct evidence (for example, a myopic witness who only got a glimpse of an object, which many subjects defined as “circumstantial”), and then given California’s “direct vs. circumstantial”

³⁰⁴ See Barbara Bergman et al., Wharton’s Criminal Evidence § 1:8 (15th ed. 2010) (collecting instructions distinguishing “direct” and “circumstantial” evidence).

³⁰⁵ Conn. Cr. Jury Instr. 2.4-1 (2008).

³⁰⁶ Cal. Crim. J. Instr. No. 223 (2006).

³⁰⁷ Conn. Cr. Jury Instr. 2.4-1 (2008), <http://www.jud.ct.gov/ji/criminal/part2/2.4-1.htm>. See Cal. Crim. J. Instr. No. 223 (2006) (“Both direct and circumstantial evidence are acceptable types of evidence to prove or disprove the elements of a charge . . . , and neither is necessarily more reliable than the other [or] entitled to any greater weight than the other.”); Greenstein, *supra* note 203, at 1803 nn.6, 7 (citing federal and state case law).

³⁰⁸ Tiersma & Curtis, *supra* note 85, at 257 (presenting study results indicating that all “direct vs. circumstantial” instructions risk reinforcing the “popular misconception that circumstantial evidence is weak”).

³⁰⁹ *Id.* at 253, 256 (finding that plain-language as well as jargon-laden instructions distinguishing “direct” and “circumstantial” evidence, even those rejecting any legal preference between the two, consistently lead jurors to associate direct evidence with stronger, more reliable evidence and circumstantial evidence with weaker evidence).

instruction, reached the wrong verdict sixty to nearly ninety percent of time.³¹⁰ Subjects incorrectly defined evidence as direct or circumstantial from 38 percent of the time (when given instructions like Connecticut’s and California’s) to 49 percent of the time (when given old-fashioned, jargon filled instructions).³¹¹

Worse, a substantial minority of American jurisdictions allows or requires jurors to privilege “direct” over “circumstantial” evidence. Some instructions define direct evidence as the norm and circumstantial evidence as something the law tolerates when “it is not . . . possible to ascertain the truth by [direct] evidence.”³¹² Others give lengthy directions about how to decide whether or not to accept circumstantial evidence as proof.³¹³ Still others never say that circumstantial evidence can be as weighty as direct evidence.³¹⁴

No jurisdiction tells jurors the truth, that *all* evidence is indirect and circumstantial and that all evidence of *identity*, including eyewitness identifications and confessions, gains strength through the aggregation of “circumstantial” matches between the defendant and what is known about the crime or criminal.³¹⁵ Instead, every criminal case arises with the prospect—and those decided at trial often end with the reality—of an authoritative legal statement nudging jurors to put more stock in the norm of “big” evidence, which seems to portray actual, “unique” events or traits of the perpetrator, and to put less stock in evidence that presents itself as only a building block, requiring acts of construction or aggregation to make something out of them. In this way the law marches deciders into the uniqueness fallacy for “big” evidence and the irrelevance fallacy for “small” non-matches.

2. *Rules Regulating Police Investigations.* — From a heuristic perspective, the worst thing investigators, forensic analysts and prosecutors can do in a whodunit case is to settle on a scenario about who did what to whom before all potential clues and witnesses have been queried, with results preserved

³¹⁰ Id. at 251-56, 259-61.

³¹¹ Id. at 250.

³¹² Pennsylvania Suggested Standard Criminal Jury Instructions § 7.02A (2010).

³¹³ See Greenstein, *supra* note 203, at 1803-04 (citing examples).

³¹⁴ See id. at 1803-04 & n.7 (citing cases).

³¹⁵ See *supra* notes 28, 77-104 and accompanying text.

and analyzed. Any such scenario triggers the unwarranted confidence associated with the simulation bias and the dangerously single-minded diversion or diminution of effort associated with a disposition to confirm what already is thought to be true. Some scenarios—ones focused on suspects with prior records, for example—also trigger the representativeness bias. By inflating the numerator, the simulation, confirmation and representativeness biases in turn ignite the uniqueness fallacy as to the suspect and irrelevance fallacy as to non-exclusionary non-matches. What then does the law do to extend evidence collection and deter scenario-formation, sequester those functions so the latter doesn't taint the former, and propagate competing scenarios?

The answer is not much. At the investigation stage, police and prosecutors are largely immune from outside supervision and scrutiny. Their key decisions—whether, when and how to investigate a crime, confiscate and document potentially evidential traces, interview potential witnesses, develop one or more scenarios and investigate one or more suspects, conduct forensic analyses, inform forensic technicians of or insulate them from police theories about who is to blame, bring the case to the district attorney to prosecute—are either entirely within the discretion of law enforcement³¹⁶ or subject only to a weak requirement of probable cause.³¹⁷ Few of the city and county departments that make these decisions have guidelines on these matters, and what guidance exists is rarely public or enforceable. State courts and attorneys general typically honor local discretion and forbear systematic data-collection or monitoring. Prosecutors have absolute immunity from damages for decisions to prosecute, including decisions based on faulty police investigations.³¹⁸

³¹⁶ See, e.g., *Wayte v. United States*, 470 U.S. 598, 607 (1985) (“[S]o long as the prosecutor has probable cause to believe that the accused committed an offense . . . , the decision whether or not to prosecute, and what charge to file or bring before a grand jury, generally rests entirely in his discretion.” (citation omitted)); *United States v. Cox* 342 F.2d 167 172 (1965) (similar).

³¹⁷ See, e.g., *Virginia v. Moore*, 553 U.S. 164, 171 (2008) (“[W]hen an officer has probable cause to believe a person committed even a minor crime in his presence, the balancing of private and public interests is not in doubt. The arrest is constitutionally reasonable.”).

³¹⁸ See, e.g., *Connick v. Thompson*, 131 S. Ct. 1350, 1366 (2011) (awarding prosecutor immunity from damages for failing to inform falsely convicted defendant of a pretrial forensic test showing that the killer's blood type excluded the defendant who spent years on death row); *Van de Kamp v. Goldstein*, 129 S. Ct. 855, 864 (2009) (granting absolute immunity to district attorney responsible for break-downs in supervision, training and information management that led to failure to disclose exculpatory information against plaintiff, who spent twenty-four years in prison for a murder he did not commit); *Joseph v. Yocum*, 53 Fed. App. 1, 3 (10th Cir.2002) (prosecutorial immunity applies to “decision to prosecute, even based on an allegedly inadequate police investigation, and the decision whether and when to dismiss the charges against plaintiff”); *Schrob v. Catterson*, 948 F.2d 1402,

In short, criminal defendants “do not enjoy a general constitutional right to a proper or thorough investigation of the offense with which they are charged.”³¹⁹ Except for “chain of custody” rules, which apply only when the state confiscates and introduces evidence at trial,³²⁰ no affirmative duties or systematic practices assure the reliable collection of evidence, interviewing of witnesses, scenario-development or suspect identification. Under the Supreme Court decision in *Youngblood v. Arizona*, officers’ negligent or reckless failure to collect or accidental destruction of evidence does not violate due process or oblige jurors to draw adverse inferences against the state.³²¹ By withholding redress absent outright bad faith, the *Youngblood* “standard imposes an almost insurmountable burden upon [an] accused” harmed by law enforcement misfeasance³²² and allowed justice to miscarry in *Youngblood*’s own case. After serving a lengthy prison term and facing prosecution for failing to register as a sex offender, *Youngblood*, through counsel, found a semen swab from the victim that could be tested using modern DNA analysis. It excluded *Youngblood* as the perpetrator.³²³ Federal and most state courts continue to apply *Youngblood*.³²⁴ Even the handful of states that have modified it only give judges discretion to instruct jurors that they may draw adverse inferences against the state if, for no good reason, officials destroyed or failed to secure important forensic evidence and if the state’s case is otherwise weak.³²⁵

1411 (3d Cir.1991) (“A prosecutor’s alleged failure to properly investigate before initiating a prosecution is . . . within the scope of absolute immunity.”).

³¹⁹ *Murphy v. Quarterman*, Nos. 3-08-CV-0826-K, 3:08-CV-1073-K 2008 WL 4899229, at *2 n.3 (N.D. Tex. Nov. 11, 2008) (quoting *Riley v. Quarterman*, 2008 WL 4425366 at *7 (S.D. Tex. Sept.24, 2008)).

³²⁰ See *Lempert et al.*, supra note 81, at 1216-21 (discussing chain of custody).

³²¹ *Youngblood v. Arizona*, 488 U.S. 51, 58 (1988) (holding that state destruction of evidence possibly favorable to the defense did not violate due process absent bad faith). Extending *Youngblood* to the failure to collect evidence are, e.g., *State v. Krosch* 642 N.W.2d 713, 719 (Minn. 2002); *State v. Ware*, 881 P.2d 679, 684 (N.M. 1994).

³²² *Norman Bay, Old Blood, Bad Blood, and Young Blood: Due Process, Lost Evidence, and the Limits of Bad Faith*, 86 Wash. U.L. Rev. 241, 247 (2008).

³²³ See, e.g., *id.* at 243; Peter Neufeld, Symposium: Serenity Now or Insanity Later?: The Impact of Post-Conviction DNA Testing On The Criminal Justice System, 35 New Eng. L. Rev. 639, 646 (2001).

³²⁴ See, e.g., *Cost v. State*, 10 A.3d 184, 193 (Md. 2010) (citing cases); *People v. Al-Azawi*, 2010 WL 4026107, at *4 (Mich. App. Oct. 14, 2010) (“A defendant is entitled to an adverse inference instruction regarding the loss of potentially exculpatory evidence only upon a showing of bad faith.”); *Bay*, supra note 322, at 274, 287 & nn.17, 364 (citing statutes).

³²⁵ See *Cost v. State*, 10 A.3d at 189 n.3, 192-93, 196-97 (citing cases and adopting rule allowing adverse-inferences if state “destroyed highly relevant evidence in its custody that it normally would have retained and submitted to forensic examination” and if the state’s case left some “doubt”); *Commonwealth v. Kee*, 870 N.E.2d 57, 62-66 (Mass. 2007) (“[W]here evidence has been lost or destroyed, it may be appropriate to instruct the jury that they may, but need not, draw an inference against the Commonwealth.”); *Bay*, supra note 322, at 287-89 & nn.373-74 (citing cases from ten allowing adverse-inference instructions).

Neither the accused nor jurors even have a right to know about defects in the state's investigation. Courts rarely find *Brady* or other violations when police or prosecutors fail to reveal defects,³²⁶ and absent bad faith, do not treat a failure to discover or collect evidence as a *Brady* violation in its own right.³²⁷ In lieu of enforcing affirmative duties to collect and test evidence or inform jurors of what was missed, American courts limit criminal defendants to cross-examining police witnesses on what they did and introducing evidence the defense has found that the state could have, but did not, obtain.³²⁸ If the defense manages to expose weaknesses in the state's investigation, it may use them only to attack the reliability of evidence the state *did* present and argue that the state's evidence leaves a reasonable doubt as to guilt.³²⁹ With the sole exception of Massachusetts, jurisdictions do not recognize an inadequate-investigation defense or require judges to instruct jurors that they may treat inadequacies in the state's investigation as sufficient in themselves to establish reasonable doubt.³³⁰

Several states have also recently adopted laws requiring preservation of DNA evidence. See *id.* at 246-47 & nn.15-17, 284-85 & nn.350-58 (citing cases and statutes).

³²⁶ See, e.g., *Woodruff v. State*, 608 N.W.2d 881, 888 (Minn. 2000) (finding no *Brady* violation despite state's failure to disclose evidence supporting defendant's inadequate-investigation claim, including officers' threats to witnesses and failure to secure the scene, collect evidence, check alibis of other suspects or turn over key reports to the prosecutor); *Guerrier v. State*, No. 54016 2010 WL 3463355, at *2 (Nev. May 07, 2010) (holding that failure to disclose police report did not violate *Brady* though it supported defendant's inadequate-investigation claims). But see *Workman v. Commonwealth*, 636 S.E.2d 368, 375-76 (Va. 2006) (finding that failure to disclose evidence supporting inadequate-investigation claim violated *Brady*).

³²⁷ See, e.g., *State v. Ware*, 881 P.2d 679, 684 (N.M. 1994) (surveying cases and holding that only evidence missed in bad faith can violate *Brady* rule).

³²⁸ See, e.g., *Harrington v. Richter*, 131 S. Ct. 770, 782 (2011) (noting that defense at trial "stressed deficiencies in the [police and forensic] investigation"); *Crawford v. Head*, 311 F.3d 1288, 1302 (11th Cir. 2002) (noting inadequate-investigation claims); *Moreno v. State*, No. A-10041 2010 WL 5021778, at *3 (Alaska App. Dec. 8, 2010) (finding that defense had fair opportunity to criticize the police investigation); *State v. Martinez*, 2010 WL 3928631, at *8 (Ariz. App. Sept. 30, 2010) (noting inadequate-investigation claims); *People v. Corbett*, 2011 WL 18733, at *14 (Cal. App. Jan. 5, 2011) (same); *State v. Jamison*, 742 A.2d 1272, 1274 (Conn. App. 1999) (same); *State v. Houston*, 191 P.3d 363 (Kan. App. 2008) (allowing cross-examine of police officers about quality of investigation); *State v. Cosey*, 779 So.2d 675, 684 (La. 2000) (noting inadequate-investigation claims); *People v. Hill*, 2010 WL 539841, at *1 (Mich. App. Feb. 16, 2010) (noting criticism of police for failing to search for fingerprints); *Fullbright v. State*, 895 A.2d 1088, 1195 (Md. App. 2006) (noting inadequate-investigation claims); *King v. State*, 47 So.3d 658, 662 (Miss. 2010) (same); *State v. Campobasso*, 123 Wash. App. 1041, *1 (Wash. App. 2004) (same). See also *Blankenship v. Hall*, 542 F.3d 1253, 1258 (11th Cir. 2008) (rejecting ineffective assistance of counsel claim based on failure to present convincing insufficient-investigation defense); *State v. Herring*, 762 N.E.2d 940, 958 (Ohio 2002) (similar).

³²⁹ See, e.g., *Morris v. State*, 927 So.2d 744, 748 (Miss. 2006) ("[T]he sufficiency or insufficiency of a police investigation simply goes to the weight of the evidence, and it is for a jury to decide what to believe." (citation omitted)); *Lane v. State*, 2009 WL 36502, at *5 (Tex. Crim. App. Jan. 8, 2009) ("The lack of physical or forensic evidence is merely a factor for the jury to consider in weighing the evidence.").

³³⁰ *Commonwealth v. Bowden*, 399 N.E.2d 482, 491 (Mass. 1980) (recognizing defense based on "fact that certain tests were not conducted or certain police procedures not followed [that] could raise a reasonable doubt as to the defendant's guilt in the minds of the jurors"). See *Commonwealth v. Silva-Santiago*, 906 N.E.2d, 299, 315 (Mass. 2009) ("Information regarding a third-party culprit, whose existence was known to the police but whose potential involvement was never investigated, may be admissible under a[n] inadequate police investigation defense . . ."); *Commonwealth v. Tolan*, 904 N.E.2d 397, 412-13 (Mass. 2009) ("The *Bowden* instruction permits jurors to consider evidence (actually presented) of police failure to take certain investigatory steps . . .

Even in Massachusetts, officers have no duty to gather exculpatory evidence.³³¹ Trial judges are simply allowed, but not required,³³² to invite jurors to consider whether omissions in the state's investigation "tend to affect the quality, reliability or credibility of the [state's] evidence."³³³ Even that limited invitation applies only if the jury finds that the omitted tests or actions (1) were standard procedure, (2) "could reasonably have been expected to lead to significant evidence of the defendant's guilt or innocence" and (3) were omitted unreasonably.³³⁴

Except perhaps in Massachusetts, therefore, applicable legal rules do little to assure the competence and thoroughness of crime scene and witness investigations, particularly in exposing non-exclusionary non-matches, and do nothing to dampen premature scenario development or selective evidence-gathering designed to confirm scenarios. In many cases, the defense won't even know about problems plaguing police investigations, and rarely will it know about non-exclusionary non-matches the police missed. The best the defense can hope to do is use cross-examination to expose documentable reasons to fear that non-matches were missed and hope the jury—unaided by instructions—will dial down its heuristically inflated confidence in the certainty of the state's unique-seeming evidence and overcome its heuristically induced blindness to individually "small," if conjointly powerful, bases for reasonable doubt.

3. *Rules Regulating Discovery.* — Even when the state obtains evidence of non-exclusionary non-matches, discovery rules invite it to withhold the information from the defense on the ground that non-matches are individually, thus collectively, irrelevant. This is particularly true after officials have committed themselves to an inculpatory scenario built around "big" evidence. In essence, the *Brady* rule

[and] such failures alone may be sufficient to create a reasonable doubt of the defendant's guilt."); Mass. Model Instr. Omissions in Police Investigations 3.740 (2009).

³³¹ See *Commonwealth v. Martinez*, 769 N.E.2d 273, 281 (Mass. 2002) ("[T]he obligation of the authorities to investigate a crime does not translate into a jury instruction that the authorities have a duty to gather exculpatory evidence.").

³³² See, e.g., *Commonwealth v. Williams*, 790 N.E.2d 662, 670 (Mass. 2003); *Commonwealth v. Lapage*, 759 N.E.2d 300, 307 (Mass. 2001).

³³³ Mass. Model Instr. Omissions in Police Investigations 3.740 (2009).

³³⁴ *Id.*

defining the constitutional duty to disclose builds the representativeness, simulation and confirmation biases into its own operation.³³⁵

The *Brady* rule comes into play at two points in the criminal process: before trial, when officials decide whether to disclose information to the defense, and on appeal, when a reviewing court decides whether the Due Process Clause requires it to overturn a conviction obtained after officials opted to withhold information. Because the Court had the latter situation in mind when it developed the standard governing both situations, the rule's application at both stages is fraught with heuristic peril, triggering the gross undervaluation and underuse of non-exclusionary non-matches.

In *Brady v. Maryland*, the Court rejected a view of prosecutions as a poker game in which holding one's losing cards close is an acceptable strategy for success.³³⁶ The Court recognized a justice interest in state³³⁷ disclosure of evidence that is both "exculpatory," because it makes more likely than without the evidence that the defendant is innocent, and "material" in that nondisclosure had some effect on the outcome.³³⁸ The Court ruled that the evidence withheld from Brady—his co-defendant's confession discussing their relative responsibility for the killing—could not have affected the guilt determination, which it upheld, but might have affected the death sentence, which it reversed.³³⁹ Since then, the Court has refined the "materiality" standard to allow reversal of a verdict only if the reviewing court finds that there is a "reasonable probability" that "but for" the nondisclosure, the jury would have reached a verdict more favorable to the defendant.³⁴⁰

³³⁵ *Brady v. Maryland*, 373 U.S. 83, 87 (1963) ("[T]he suppression by the prosecution of evidence favorable to an accused . . . violates due process where the evidence is material either to guilt or to punishment . . ."). See *United States v. Bagley*, 473 U.S. 667, 682 (1985) (defining undisclosed evidence as material "if there is a reasonable probability that, had the evidence been disclosed . . ., the result of the proceeding would have been different").

³³⁶ *Brady v. Maryland*, 373 U.S. at 86-88. See *Williams v. Florida*, 399 U.S. 78, 82 (1970) ("The adversary system of trial . . . is not yet a poker game in which players enjoy an absolute right always to conceal their cards until played."); William J. Brennan, Jr., *The Criminal Prosecution: Sporting Event or Quest for Truth?*, 1963 Wash. U. L.Q. 279, 292.

³³⁷ See *Kyles v. Whitley*, 514 U.S. 419, 437-38 (1995) ("[T]he individual prosecutor has a duty to learn of [and disclose] any favorable evidence known to the others acting on the government's behalf in the case, including the police.")

³³⁸ "Society wins not only when the guilty are convicted but when criminal trials are fair A prosecution that withholds evidence . . . which, if made available, would tend to exculpate him or reduce the penalty . . . casts the prosecutor in the role of an architect of a proceeding that does not comport with . . . justice . . ." *Brady v. Maryland*, 373 U.S. at 87-88.

³³⁹ *Id.* at 84.

³⁴⁰ *United States v. Bagley*, 473 U.S. 667, 682 (1985) (quoted *supra* note 335).

A problem with the *Brady* rule is that it does not tell prosecutors what they are expected to disclose *before trial*—for example, “turn over all evidence favorable to the defense, so defense counsel and the jury can decide what is valuable,” or “show all exculpatory evidence to the judge who will decide what the defense and jury should see,” or even “disclose all ‘material’ evidence, that a reasonable trier of fact would consider important.” Rather, the rule tells officials only to keep in mind what will happen *on appeal* if a court finds a reasonable probability that something they did not turn over would have changed the outcome of the trial to come. Paraphrasing Professor Burke, *Brady* tells the “virtuous prosecutor” trying to decide whether to reveal potential loser cards to “do whatever you want as long as you don’t get reversed.”³⁴¹

Bizarrely, that is, *Brady* requires the conscientious prosecutor to decide whether to withhold exculpatory evidence by predicting the *future* determination an appellate court *might* make about whether the *past* course of history would likely have changed if the prosecutor had made a different prediction of the court’s *future* ruling.³⁴² Worse, by far the most likely future outcome of all is that the appellate court will make no ruling, in which case the prosecutor has nothing to worry about because her only worry is reversal. A “no decision” is as good as an affirmance, and one of those outcomes will almost certainly occur unless five conditions are met: (1) the defendant is not offered or declines a plea bargain, which usually forestalls an appeal; (2) the trial defendant is convicted; (3) the defendant learns that the prosecutor withheld exculpatory evidence; (4) the defendant appeals on that ground; (5) the reviewing court finds the withheld evidence “material.”

Condition (1) is rare, to begin with, because over 90 percent of prosecutions end in plea bargains.³⁴³ Indeed, the prosecutor at least marginally increases the chance of a plea by withholding exculpatory evidence that might embolden the defendant to go to trial. Condition (3) also is rare, because

³⁴¹ Alafair Burke, *Brady’s Brainteaser: The Accidental Prosecutor and Cognitive Bias*, 57 Case W. Res. L. Rev. 575, 576 (2007).

³⁴² See Burke, *Prosecutorial Decision Making*, supra note 208, at 1609-12 (“[A]pplying the [*Brady*] standard prior to trial requires that prosecutors engage in a bizarre kind of anticipatory hindsight review,” inviting confirmation bias at multiple points); Bennett L. Gershman, *Litigating Brady: Games Prosecutors Play*, 57 Case W. Res. L. Rev. 531, 558 (2007) (criticizing post hoc evaluation *Brady* rule requires).

³⁴³ See Ronald J. Allen et al., *Comprehensive Criminal Procedure* 1161 (2d. ed. 2005) (“Of felony convictions nationwide, 94 percent are obtained by guilty plea.”).

the prosecution controls the evidence it withheld, making it hard for defendants to discover. Finally, reversals of criminal verdicts on any, much less on *Brady*, grounds (condition (5)) occur in only a small percentage of the minuscule subset of cases meeting the other conditions.³⁴⁴ Any appellate determination, especially a reversal, is unlikely, making it rational for conscientious prosecutors to err on the side of withholding evidence. As Justice Marshall noted, the “materiality” standard gives prosecutors an “impossible task” of predicting “whether a certain piece of information will have a significant impact on the trial” based on “speculat[ion] . . . without foundation” about “what strategy the defense will pursue or what evidence” it will present; it “invites a prosecutor, whose interests are conflicting, to gamble . . . that evidence will later turn out not to [be held to] have been potentially dispositive.”³⁴⁵

Still worse, when prosecutors predict, speculate and play the odds, heuristic biases assure that few pieces of exculpatory evidence—particularly non-exclusionary non-matches—will strike them as having to be disclosed.³⁴⁶ Before requiring disclosure, *Brady* obliges the prosecutor to imagine her way through several future events: She goes to trial and presents the inculpatory scenario she and police officers have developed, a scenario often based in part on the defendant’s prior criminal record. She lays out the scenario, say, through eyewitness testimony identifying the defendant or his own admissions. The jury finds the state’s theory true beyond a reasonable doubt and convicts. The appellate court considers whether the withheld evidence would have made a difference after drawing the strongest inferences in favor of the state that its trial evidence allows.³⁴⁷

These mental steps all but assure that prosecutors will succumb to excessive confidence spurred by:

³⁴⁴ See, e.g., Nancy J. King & Joseph L. Hoffmann, Habeas for the Twenty-First Century: Uses, Abuses, and the Future of the Great Writ 197 n.45 (2011) (“[R]eversal rates in federal habeas, except in capital cases, probably never reached double digits, even during the heyday of the Warren Court”); Richard Faust, et al. The Great Writ in Action: Empirical Light on the Federal Habeas Corpus Debate, 18 N.Y.U. Rev. L. & Soc. Change 637, 677-81 (finding 3% reversal rate in 1970s federal habeas corpus cases in Southern District of New York).

³⁴⁵ United States v. Bagley, 473 U.S. at 701 (Marshall, J., dissenting).

³⁴⁶ See Burke, *Brady’s Brainteaser*, supra note 341, at 576-77 (arguing that cognitive biases invite “prosecutors to systematically undervalue the materiality of evidence”).

³⁴⁷ The *Brady* rule requires reviewing courts to make an “independent examination of the record,” United States v. Bagley, 473 U.S. at 679 n.8, and forbear reversing unless withheld evidence is strong enough to alter the jury’s “beyond a reasonable doubt” confidence in the evidence it heard, see *Cone v. Bell*, 129 S. Ct. 1769, 1783 (U.S. 2009); *Youngblood v. West Virginia*, 547 U.S. 867, 870 (2007).

- The *simulation bias* thrice over. In bringing charges, the prosecutor adopted the scenario police offered as the basis for prosecution. *Brady* then requires her to imagine that she successfully presented the scenario to the jury. Then she must imagine an appellate court reviewing the scenario and indulging every presumption in its favor.³⁴⁸
- The *representativeness* bias. The defendant's prior record and the scenario itself cast the defendant as someone who resembles a criminal, obscuring the base rate of other possible suspects.³⁴⁹
- The *confirmation bias*. In imagining an appellate decision as to whether the withheld evidence would have changed the outcome, she will remember the strong parts of her case and weaknesses in the evidence under consideration.³⁵⁰
- An *outcome bias* twice over.³⁵¹ Both the decision to prosecute and the imagined guilty verdict will lend a false validity to the scenario that triggered the outcome.
- The *uniqueness* and *irrelevance fallacies*. By leading decision makers to assign a high numerator value to inculpatory implications of the state's evidence and assume innocent explanations for non-exclusionary non-matches, the above biases will create a false certainty in the uniqueness of the former and irrelevance of latter.³⁵²

Finally, if the prosecutor is thinking clearly, she will realize that the appellate court itself will be prey to these biases, particularly the outcome bias based on the jury verdict, further diminishing the microscopic probability of reversal.

Of course, if the exculpatory evidence the prosecutor possesses is “big”—a credible confession by someone besides the defendant or non-matching DNA on a rape victim's vaginal swab—we expect

³⁴⁸ See supra notes 197-207 and accompanying text.

³⁴⁹ See supra notes 185-191 and accompanying text.

³⁵⁰ See Burke, *Brady's Brainteaser*, supra note 341, at 578-80 (noting that, via the confirmation bias, imagining the trial absent the evidence in question will lead the prosecutor “to recall facts that support” and ignore “facts that might undermine her existing belief in the defendant's guilt” and to “scrutinize the potentially exculpatory evidence for flaws”); supra notes 208-214 and accompanying text.

³⁵¹ Outcome bias leads decision makers to treat the very fact of an outcome as retroactively validating poor decisions producing the outcome. See, e.g., Jonathan Baron & John C. Hershey, Outcome Bias in Decision Evaluation, 54 J. Personality & Soc. Psych. 569, 570 (1988) (describing studies indicating that “people take outcomes into account in a way that is irrelevant to the true quality of the decision”); Kahneman, Thinking, Fast, supra note 184, at 203-04.

³⁵² See supra notes 235-269 and accompanying text.

materiality to be obvious and the prosecutor to disclose, heuristic biases notwithstanding. If, however, the exculpatory evidence is “small”—a non-matching stray button or a detail in an eyewitness’ initial description to police—the biases the *Brady* rule invites will create a myriad of reasons why the prosecutor will rate the after-the-fact effect of each non-match by itself, and the aggregate effect of them all, as low or nil.

4. *Rules Limiting Evidence of a Third Party’s Guilt.* — As we note above, the drama that plays out in adversarial trials is not a “*Whodunit?*,” but a “*Did Smith Do It?*,” triggering representativeness, simulation and confirmation bias from the moment the case is styled “*People v. Smith.*”³⁵³ If Smith could present evidence establishing a counter-scenario, with a different suspect in the role of villain, heuristic problems could be allayed.³⁵⁴ But realities of our criminal justice system and rules limiting evidence of a third party’s guilt make it difficult to mount an effective “I didn’t do it, but I’ll tell you who did” defense.

Realistically, few defendants, especially innocent ones disconnected from the events charged, have meaningful access to evidence implicating third-party suspects. Separating defendants from evidence of other suspects are law enforcement’s monopoly over the crime scene and associated witnesses as well as grand jury proceedings,³⁵⁵ the absence of a duty to conduct police investigations competently or face consequences,³⁵⁶ heuristic and legal limits on officials’ obligation to turn over leads that do not support the state’s case,³⁵⁷ chronic underfunding of indigent criminal defense investigations,³⁵⁸ defense counsel’s disposition to give priority in deploying limited resources to poking holes in the state’s

³⁵³ See supra notes 39, 301-302 and accompanying text.

³⁵⁴ See Heller, supra note 18, at 290 (showing that “priming” with alternative, exculpatory scenarios can overcome unfair heuristic advantages accompanying the state’s “direct” evidence).

³⁵⁵ See supra notes 283-300 and accompanying text.

³⁵⁶ See supra notes 316-334 and accompanying text.

³⁵⁷ See supra notes 335-352 and accompanying text.

³⁵⁸ See, e.g., ABA Standing Committee on Legal Aid and Indigent Defense, *Gideon’s Broken Promise: America’s Continuing Quest for Equal Justice* 10-11, 38 (2004), available at http://www.americanbar.org/content/dam/aba/migrated/legalservices/sclaid/defender/brokenpromise/right_to_counsel_in_criminal_proceedings_fullreport.authcheckdam.pdf (Apr. 17, 2010) (discussing chronic inadequacy of resources for pretrial defense investigation).

case not constructing her own, and the disposition of most criminal cases by plea bargains aimed at minimizing the cost of investigations.³⁵⁹

Further decreasing incentives to look for evidence of a third-party's guilt are mystifying legal doctrines limiting the admissibility of such evidence at trial. The usual rule, of course, is that the prosecution bears the burden of proving identity, so any evidence the defendant presents suggesting that someone else committed the crime—an alibi, for example—obliges the state to dispel the implication beyond a reasonable doubt.³⁶⁰ Typically, the defendant can trigger this process with any evidence that is relevant and not substantially more prejudicial than probative.³⁶¹ Because an edifice of guilt or innocence may be built “brick by brick,” triggering evidence need only minutely change the probabilities of guilt or innocence and need not meet a burden of proof by itself.³⁶² The defendant's Sixth Amendment right to present defensive evidence undergirds these rules.³⁶³

But this is not how the law works when it comes to the defense that Jones, not the defendant Smith, did it. As the Supreme Court noted recently in *Holmes v. South Carolina*, most jurisdictions strictly limit the admissibility of concededly relevant evidence that implicates, or orients non-exclusionary non-matches towards, a specified alternative suspect.³⁶⁴ In these jurisdictions, such evidence is inadmissible unless it (1) establishes a “clear link” or “direct connection” between the alternate suspect

³⁵⁹ See Allen et al., *Criminal Procedure*, supra note 343, at 1186-87 (noting that resource constraints lead defense lawyers to use plea bargaining to “triage” among cases based on predictions about the relative value of additional investment of attorney time).

³⁶⁰ See, e.g., American Law Institute, *Model Penal Code* §§ 1.12(1), 1.12(2)(a), 1.13(9)(c) (2001) (providing that once evidence negating an element of the offense is offered, the state must disprove that defense beyond a reasonable doubt).

³⁶¹ See, e.g., *Fed. R. Evid.* 401-403.

³⁶² *Advisory Comm. Note to Fed. R. Evid.* 402 (1972) (“The standard of probability under the rule is ‘more * * * probable than it would be without the evidence.’ Any more stringent requirement is unworkable and unrealistic. As *McCormick* § 152, p. 317, says, ‘A brick is not a wall.’”). See *Bourjaily v. United States*, 483 U.S. 171, 179-180 (1987) (“[I]ndividual pieces of evidence, insufficient in themselves to prove a point, may in cumulation prove it.”).

³⁶³ See, e.g., *Crane v. Kentucky*, 476 U.S. 683, 691 (1986); *Chambers v. Mississippi*, 410 U.S. 284, 294, 297 (1973); John H. Blume et al., *Every Juror Wants a Story: Narrative Relevance, Third Party Guilt, and the Right to Present a Defense*, 44 *Am. Crim. L. Rev.* 1069, 1103 (2007) (“When there is credible evidence of a third party's potential guilt, then strict restrictions on admissibility of such evidence unreasonably infringe upon a criminal defendant's right to present a complete defense.”).

³⁶⁴ *Holmes v. South Carolina*, 547 U.S. 319, 327 & n.* (2006) (noting “wide” acceptance of “rules regulating the admission of evidence proffered by criminal defendants to show that someone else committed the crime” (citing cases)); *Cleveland v. State*, 91 P.3d 965, 972 (Alaska Ct. App. 2004) (“[V]irtually every state . . . require[s] some kind of preliminary evidentiary showing before allowing introduction of alternative-perpetrator evidence.” (citing cases)); David McCord, “But Perry Mason Made It Look So Easy!”: The Admissibility of Evidence Offered by a Criminal Defendant to Suggest that Someone Else is Guilty, 63 *Tenn. L. Rev.* 917, 936-38, 987 (1996) (criticizing majority rule placing a “direct connection” limitation on third-party evidence and alternative rules imposing a “capable of raising a reasonable doubt” standard that is stricter than the usual balance of probative weight and prejudice or applying the usual balance in an unusually strict manner when evaluating third-party evidence).

and the crime (the majority rule);³⁶⁵ (2) satisfies the defense's effective standard of *proof* by singlehandedly establishing a reasonable doubt;³⁶⁶ or (3) is more probative than prejudicial, reversing the usual presumption in favor of admissibility.³⁶⁷ Applying these doctrines, courts have excluded evidence that the father of the sexual-abuse victim was convicted of sexual abuse several years after the incident with which the defendant is charged;³⁶⁸ police initially arrested another man for the crime;³⁶⁹ a man (whom the trial court forbade the defendant to call at trial) was identified by a paramedic as resembling a man at the crime scene who acted suspiciously, then fled;³⁷⁰ statements police obtained from a man with a motive to commit the crime revealed an unexplained familiarity with details of the crime;³⁷¹ and another man possessed a distinctive box resembling one taken from the victim around the time of the crime.³⁷² Only when evidence implicating a third-party suspect verges on "big" evidence³⁷³—DNA that excludes the defendant and is consistent with the alternate suspect or an eyewitness identification of that suspect—is the defendant likely to be allowed to introduce the evidence.³⁷⁴

Recently, the Supreme Court overturned a South Carolina rule providing that, regardless of the strength of the evidence linking an alternative suspect to the crime, "where there is strong evidence of [the *defendant's*] guilt, especially . . . forensic evidence, the proffered evidence about a third party's

³⁶⁵ See, e.g., McCord, *supra* note 364, at 936-38 (citing cases).

³⁶⁶ See, e.g., 41 C.J.S., Homicide § 216, at 56-58 (1991) ("Evidence tending to show the commission by another person of the crime charged may be introduced by accused when it is inconsistent with, and raises a reasonable doubt of, his own guilt; but . . . evidence . . . lack[ing] such connection [is] excluded."); McCord, *supra* note 364, at 936-38 (citing cases).

³⁶⁷ See McCord, *supra* note 364, at 936-38 (citing cases).

³⁶⁸ *People v. Sparman*, 608 N.Y.S.2d 672, 673-74 (N.Y. App. Div. 1994) (excluding evidence implicating alternate suspect with greater access to child-abuse victim because such evidence "must do more than raise a mere suspicion that another person committed the crime; there must be a clear link between the third party and the crime").

³⁶⁹ *State v. Oliver*, 821 P.2d 250, 251-52 (Ariz. App. 1991) (rejecting alternative-suspect evidence absent an "inherent tendency to connect the person with the actual commission of the crime"); *People v. Owens*, 97 P.3d 227, 235 (Colo. App. 2004) (requiring a "direct connection" between an alternative perpetrator and the crime).

³⁷⁰ *State v. Woodfield*, 659 P.2d 1006, 1008 (Or. App. 1983).

³⁷¹ *State v. Gilman*, 608 A.2d 660, 663 (Vt. 1992) (barring alternative-suspect evidence unless it incriminates the third party "directly").

³⁷² *State v. McNeill*, 392 S.E.2d 78, 83-84 (N.C. 1990) (barring alternate-suspect evidence that is "not inconsistent with the defendant's own guilt"). See McCord, *supra* note 364, at 950-51 & n.167 (discussing cases).

³⁷³ See, e.g., *State v. McNeill*, 392 S.E.2d at 84 (discussed *supra* note 372).

³⁷⁴ See, e.g., *Holmes v. South Carolina*, 547 U.S. 319, 323, 330-31 (2006) (discussed *infra* notes 375-377 and accompanying text); *United States v. Armstrong*, 621 F.2d 951, 953 (9th Cir. 1980) (requiring admission of evidence that alternate suspect had bait money from robbery); *United States v. Robinson*, 544 F.2d 110, 112-13 (2d Cir. 1976) (reversing conviction because trial court excluded evidence that a correctional officer identified someone other than the defendant as a bank robber recorded on a surveillance videotape); *Joyner v. State*, 678 N.E.2d 386, 388-89 (Ind. 1997) (requiring admission of mitochondrial DNA evidence of hair found on the victim that excluded the defendant and matched the alternate suspect); McCord, *supra* note 364, at 951 & n.168 (citing similar cases).

alleged guilt' may (or perhaps must) be excluded."³⁷⁵ Despite this ruling, the Court seemed to approve a rule other states apply, that "evidence offered by accused as to the commission of the crime by another person must be limited to such facts as are inconsistent with his own guilt, and . . . raise a reasonable inference or presumption as to his own innocence."³⁷⁶ By emphasizing the "big" character of the third-party evidence before it—the alternative suspect was near the rape-murder victim's home at the time of the crime and told four witnesses that he had committed the crime or that Holmes was innocent—the Court gave no solace to defendants relying on even a constellation of "small" non-exclusionary non-matches as to themselves that match an alternative suspect.³⁷⁷ Rather, *Holmes* seems to recognize a right only to fight fire with fire—to oppose the state's "big" evidence implicating the defendant with "big" evidence implicating a third party—while allowing states to forbid defendants to fight fire with even a torrent of "small" non-matches. While mitigating the uniqueness fallacy, *Holmes* bolsters the irrelevance fallacy.

5. *Rules Limiting Statistical Evidence.* — In *People v. Collins*,³⁷⁸ the California Supreme Court overturned a conviction based on a fumbling, possibly invidious, effort by a mathematics instructor to show how improbable it was that there were two interracial couples in Los Angeles who, like the defendant couple, matched the perpetrators' description. As we note above, however, rather than limit its reversal to the expert's manifold technical mistakes,³⁷⁹ the court ruled that even *properly* applied statistical techniques have no place in adjudicating identity. The court supposed that, when "[c]onfronted with an equation which purports to yield a numerical index of probable guilt, few juries could resist the temptation to accord disproportionate weight to that index."³⁸⁰ It further assumed that "no mathematical equation can prove beyond a reasonable doubt (1) that the guilty couple in fact possessed the characteristics described by the People's witnesses, or even (2) that only one couple possessing those

³⁷⁵ *Holmes v. South Carolina*, 547 U.S. at 329.

³⁷⁶ *Id.* at 323-34, 328.

³⁷⁷ *Id.* at 323, 329 (rejecting South Carolina rule because it "does not focus on the probative value" of the defendant's "evidence of third-party guilt").

³⁷⁸ 438 P.2d 33 (Cal. 1968).

³⁷⁹ *Id.* at 39 (discussed *supra* notes 58-62 and accompanying text).

³⁸⁰ *Id.* at 40.

distinctive characteristics could be found in the entire Los Angeles area.”³⁸¹ What we now know about heuristic bias—particularly the certainty effect—undermines both assumptions. Contrary to the first, human decision makers’ quest for certainty leads them systematically to under-, not over-, value explicit probabilities.³⁸² Worse, the claim that the aggregate of lesser probabilities can never be high enough to prove guilt (or even, it seems, reasonable doubt) simply codifies the certainty fallacy as the law of the land. In an understandable effort to head off a single miscarriage of justice, the decision increased the risk of many more. It unnecessarily barred a valuable method of deterring decision makers from overvaluing “big” evidence of guilt that dangerously masquerades as “unique” identifiers—the eyewitness identification of Carlos DeLuna, for example, and the fingerprint “match” of Brandon Mayfield—and from undervaluing non-exclusionary non-matches (including in those cases) that the certainty bias and the court’s ruling render invisible and effectively irrelevant.³⁸³

Despite these flaws, *Collins*’ reasoning as elucidated by Professor Tribe is accepted by many courts as a sufficient reason to bar Bayesian and allied statistical analysis.³⁸⁴ Illustrative is the Minnesota Supreme Court’s blanket ban on even admittedly sound probabilistic estimates of the frequency of

³⁸¹ *Id.*

³⁸² See *supra* notes 218-234 and accompanying text.

³⁸³ See *supra* notes 10-26, 270-276, 285-297 and accompanying text.

³⁸⁴ See, e.g., *United States v. Gwaltney*, 790 F.2d 1378, 1383 (9th Cir. 1986) (citing *Collins* on standard for admitting probabilistic evidence); *DiGiacomo v. Franzen*, 680 F.2d 515, 518 (7th Cir. 1982) (citing *Collins* to support view that probabilities confuse jurors); *Hart v. Secretary*, 60 Fed. Cl. 598, 606 (2004) (citing *Collins* in invalidating master’s reliance on probabilistic analysis to rule out a vaccine as the cause of an infant’s death); *United States v. Yee*, 134 F.R.D. 161, 211 (N.D. Ohio 1991) (citing *Collins* in ruling probabilistic testimony inadmissible); *People v. Cella*, 188 Cal. Rptr. 675, 684 (Ct. App. 1983) (allowing “mathematical probabilities” only if “all crucial variables [are] quantified exactly”); *Davis v. State*, 476 N.E.2d 127, 134 (Ind. App. 1985) (citing *Collins* in barring probabilistic evidence absent “empirical scientific data”); *Wilson v. State*, 803 A.2d 1034, 1044-46 (Md. 2002) (barring multiplication rule in assessing likelihood of two infants in same family dying of Sudden Infant Death Syndrome (SIDS)); *Commonwealth v. Drayton*, 434 N.E.2d 997, 1005 (Mass 1982) (disallowing fingerprint expert’s estimate of “statistical probability that prints with twelve points of similarity could be made by two different people”); *Buchanan v. State*, 69 P.3d 694, 709 (Nev. 2003) (Rose, J., concurring) (citing “landmark” *Collins* decision in invalidating statistics as deficient basis for concluding that intentional asphyxiation, not SIDS, caused child’s death); *Pearson v. State*, 811 P.2d 704, 707-08 (Wyo. 1991) (citing *Collins* in excluding probabilistic evidence); *R v Adams*, 1 Crim. App. R. 377, 384 (1997) (“To introduce Bayes’ Theorem, or any similar method, into a criminal trial plunges the jury into inappropriate and unnecessary realms of theory and complexity deflecting them from their proper task.” (discussed *supra* notes 43-57 and accompanying text)); Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 *J. of Leg. Stud.* 107, 135 (2007) (arguing that Bayesian analysis is of little value in reaching reliable trial outcomes); Richard O. Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof, in Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism* 61-63 (Peter Tillers & Eric Green eds. 1988) (“Among legal academics, it is generally agreed that Tribe won this particular debate [about the value of Bayesian analysis in court].”). For more sympathetic views of the Bayesian and similar analysis, see, e.g., Lempert, *supra* at 61-63; Nance, *supra* note 105, at 1595-616; Posner, *supra* note 28, at 1508 (quoted *supra* note 28).

identifying traits in the population as an aid in adjudicating identity in criminal cases.³⁸⁵ Even when the court created an exception for DNA evidence,³⁸⁶ its recognition that this most “unique”-seeming of evidence in fact derives its strength from the use of the multiplication rule to aggregate multiple individually unimpressive probabilities did not lead it to relax its general ban on statistical aggregation.³⁸⁷ In Minnesota, as elsewhere, courts continue to resist the three steps needed to implement a Bayesian solution in criminal cases in which identity is in doubt: letting the parties (1) offer evidence of the number of possible suspects (the base rate), (2) establish the frequency of certain traits in the population (for use in estimating the numerator and denominator), and (3) present expert testimony or secure instructions to guide jurors in proper statistical analysis.

To be sure, gaps have appeared in the *Collins* rule. Most courts now allow evidence of the frequency of DNA and other “forensic” traits in the population and thus of the probability of a random match between forensic evidence and a suspect (step 2 above).³⁸⁸ Nearly all allow testimony relying on the multiplication rule in inculpatory DNA cases (step 3). In criminal and civil paternity cases, some courts have gone further, letting experts use DNA tests to estimate the Bayesian likelihood-ratio numerator and denominator associated with genetic matches between a child and a putative parent (step 2), experimenting with ways to help jurors estimate the prior odds of paternity based on non-scientific evidence (step 1), and inviting jurors to decide based on the Bayesian formula and charts with Bayesian

³⁸⁵ See *State v. Joon Kyu Kim*, 398 N.W.2d 544, 548 (Minn. 1987) (excluding expert testimony that semen found in victim’s body and on a bed was consistent with Kim’s blood type and that 96.4 percent of males in Twin Cities could be excluded as possible sources of the semen); *State v. Boyd*, 331 N.W.2d 480, 483 (Minn. 1983) (barring evidence of 99.911 percent probability of paternity based on blood tests); *State v. Carlson*, 267 N.W.2d 170, 176 (Minn. 1978) (concluding that, although probabilities offered by hair expert were methodologically sound, the predicted psychological effect on the jury was too great to allow their admission: “Testimony expressing opinions or conclusions in terms of statistical probabilities can make the uncertain seem all but proven, and suggest, by quantification, satisfaction of the requirement that guilt be established beyond a reasonable doubt.”).

³⁸⁶ See, e.g., *State v. Bloom*, 516 N.W.2d 159, 167 (Minn. 1994) (establishing “DNA exception to the rule against admission of quantitative, statistical probability evidence in criminal prosecutions to prove identity”).

³⁸⁷ See *State v. Hannon*, 703 N.W.2d 498, 508 (Minn. 2005) (extending DNA exception to multiple-source DNA samples but reaffirming “general prohibition against admission of statistical probability evidence in criminal prosecutions”).

³⁸⁸ See, e.g., *People v. Mountain*, 486 N.E.2d 802, 804 (N.Y. 1985) (reversing prior precedent and allowing evidence of blood-type matches and associated frequencies); sources cited *supra* note 105.

outcomes for different estimates of prior odds, numerator and denominator (step 3).³⁸⁹ With limited exceptions, however, courts continue to resist evidence of the frequency of *non*-forensic traces (step 2) unless the reference class used to determine the frequency “exactly” matches the facts of the case at hand—a barrier “so extreme that [it] would eliminate the use of statistical evidence” even for DNA and other forensic evidence.³⁹⁰ Likewise, uses of the multiplication rule (step 3) and Bayesian analysis (steps 1-3) are unheard of outside DNA cases,³⁹¹ and we know of no instances in which these techniques have been used to systematically aggregate the force of multiple non-exclusionary non-matches to establish a reasonable doubt. To this extent, *Collins* remains the norm, barring the most direct method of dispelling the uniqueness and irrelevance fallacies.

V. SENSIBLE REGULATION OF NON-EXCLUSIONARY NON-MATCHES AS EVIDENCE OF IDENTITY

The forces described here are hard-wired into our minds, our court system and our law. Is it realistic to think they can be changed sufficiently to make a difference? In this Part, we describe two new sets of tools for addressing the problem. The first illustrates a point made earlier: adversarial pressures can trigger innovative solutions to problems that arise in translating probabilistic into trial proofs. The second set of tools is a regulatory mechanism through which iterative procedural and legal innovation can take place in advance of certainty about the best or full range of available solutions.

A. *New Tools to Improve Lay Decision Makers' Appreciation of Aggregative Analysis*

There is a growing body of literature and practice on how to organize human behavior to circumvent heuristic biases.³⁹² Part of the literature, emanating from computer scientists and mathematicians in the UK, focuses on ways to help jurors and other criminal-process actors make

³⁸⁹ See, e.g., *In re Paternity of M.J.B.*, 425 N.W.2d 404, 408-09 (Wis. 1988) (describing different statistical techniques); sources cited *supra* note 104; *supra* notes 156 and accompanying text.

³⁹⁰ See Koehler, *When Do*, *supra* note 105, at 380-93 (citing cases; discussed *supra* note 106).

³⁹¹ See Nance, *supra* note 105, at 1612 (“Neither the likelihood ratio format nor the chart format is commonly employed at this time in criminal cases in the United States [apart from] paternity cases.”).

³⁹² See sources cited *supra* notes 277-280.

effective use of Bayesian analysis,³⁹³ including in assessing the conjoint effect of non-exclusionary non-matches.³⁹⁴ Two of these experts, Norman Fenton and Martin Neil, take as their starting point that “for most people—and this includes . . . highly intelligent barristers, judges and surgeons—any attempt to use Bayes’ Theorem is completely hopeless.”³⁹⁵ Judges and lawyers “cannot be expected to follow even the simplest instances of Bayes’ Theorem in its formulaic representation” and “simply switch-off at the sight of [the] formula.”³⁹⁶ As the authors’ experimental studies have shown, however, “it is the use of abstract probabilities and formulas, rather than the underlying concept, that act as a barrier to understanding. When the Bayesian argument is presented visually, using [graphic representations of] concrete frequencies people not only generally understand it . . . , they can construct their own simple calculations.”³⁹⁷

Consider also a well-known study revealing the proneness of experts—Harvard Medical School faculty and students—to make the same mistakes lay people do by ignoring base rate information. The experts were asked to gauge the chance that a person with a positive result on a test for a disease with a

³⁹³ See, e.g., Colin Aiken et al., A Graphical Model for the Evaluation of Cross-Transfer Evidence in DNA Profiles, 63 *Theoretical Pop. Bio.* 179 (2003); A. Biedermann & F. Taroni, Bayesian Networks and Probabilistic Reasoning about Scientific Evidence When There Is a Lack of Data, 157 *Forensic Sci. Intern’l* 163 (2006); Leda Cosmides & John Tooby, Are Humans Good Intuitive Statisticians After All?, 58 *Cognition* 1, 68-69 (1996) (citing sources); A. Philip Dawid, Bayes’s Theorem and Weighing Evidence by Juries, in *Bayes’ Theorem*, 113 *Proceed. British Acad.* 71 (2002); A. Philip Dawid & Ian W. Evett, Using a Graphical Model to Assist the Evaluation of Complicated Patterns of Evidence, *J. Forensic Sci.* 226 (1997); Donnelly, *supra* note 43; Norman Fenton & Martin Neil, Avoiding Probabilistic Reasoning Fallacies in Legal Practice Using Bayesian Networks (2008), http://www.eecs.qmul.ac.uk/~norman/papers/fenton_neil_prob_fallacies_3_July_09.pdf; Norman Fenton & Martin Neil, The “Jury Observation Fallacy” and the Use of Bayesian Networks to Present Probabilistic Legal Arguments, 36 *Math. Today* 180 (2000) (revised July 31 2007), http://www.dcs.qmul.ac.uk/researchgp/spotlight/jury_fallacy.pdf, at 2-10; Luke M. Froeb & Bruce H. Kobayashi, Naïve, Biased, Yet Bayeseian: Can Juries Interpret Selectively Produced Evidence, 12 *J. L. Econ. & Org.* 257 (1996); Amanda B. Hepler et al., Object-Oriented Graphical Representations of Complex Patterns of Evidence, 6 *L., Prob. & Risk* 275, 281 (2007) (“[Bayesian analysis] can be useful for the purely qualitative analysis of masses of evidence”); C. Jowett, *Sittin’ in the Dock with the Bayes*, *New L.J. Expert Witness Supp.* (Feb 16, 2001); Nance, *supra* note 105, at 1614-16; Nance & Morris, *supra* note 156, at 403, 421, 429 (reporting study findings that, although subjects typically undervalue forensic match evidence, use of “chart format” showing effect of different prior odds on the likelihood ratio associated with DNA evidence can substantially reduce error); Bernard Robertson & G. A. Vignaux, *Explaining Evidence Logically*, 148 *New L.J.* 159 (1998).

³⁹⁴ Donnelly, *supra* note 43, at 16 (describing questionnaire prepared by experts for both sides to help jurors with Bayesian analysis in *R v. Adams*); Fenton & Neil, *Avoiding Fallacies*, *supra* note 113, at 13 (discussing jurors’ use of Bayes’ Theorem “when there are multiple pieces of contradictory evidence and interdependencies between them”).

³⁹⁵ Fenton & Neil, *Avoiding Fallacies*, *supra* note 113, at 13. See Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 *Va. L. Rev.* 1491, 1545 (2001) (observing that “lay people make a mess” of probabilistic evidence); Tribe, *supra* note 67, at 1368 (arguing that Bayes’ Theorem is “completely opaque to all but the trained mathematician”); studies cited *supra* note 193.

³⁹⁶ Fenton & Neil, *Avoiding Fallacies*, *supra* note 113, at 13, at 2, 13.

³⁹⁷ *Id.* at 17.

prevalence of 1/1000 and a false positive rate of 5% “*actually* has the disease.” Only 18 percent of the experts gave the correct answer of less than 2 percent ($1/999 \times .95/.05 = .95/49.95 \rightarrow 1.87\%$); 45 percent ignored the base rate entirely and answered 95 percent.³⁹⁸ When Leda Cosmides and John Tooby replicated the study with Stanford undergraduates, their subjects performed even worse: 12 percent got the right answer, and 56 percent ignored the base rate.³⁹⁹ By taking a few simple steps however, the researchers made most of the error go away. First, they asked the same question of a new set of undergraduates but (1) expressed the proportions by inviting subjects to imagine large groups of people with the relevant frequencies, and (2) asked subjects to give an answer not as a probability that a single person with a positive result would have the disease but as the expected proportion of many people with positive results who actually had the disease.⁴⁰⁰ This reversed the results: 56 percent got the right answer; only 4 percent ignored the base rate.⁴⁰¹ By asking a series of “probe questions”—akin to special verdicts—that modeled key steps in correct Bayesian analysis, they hiked the number of subjects who got the right answer to 76 percent, with *none* entirely ignoring the base rate.⁴⁰² Even without probe questions, 76 percent of subjects answered correctly when references to “percentages” were omitted and frequencies were stated as “how many out of 1000 people” fell into particular categories.⁴⁰³ Finally, when the researchers created a pictorial schema of a grid of tiny boxes representing the “1000 Americans” and asked subjects to answer the probe questions by coloring in the boxes representing the relevant number of individuals, correct answers rose to 92 percent.⁴⁰⁴

³⁹⁸ W. Casscells et al., *supra* note 188, at 999-1001.

³⁹⁹ Cosmides & Tooby, *supra* note 393, at 25.

⁴⁰⁰ Subjects were told that “1 out of every 1000 Americans has the disease” and “out of every 1000 people who are perfectly healthy, 50 of them test positive for the disease (i.e., the ‘false positive’ rate is 5%)” and asked, “on average, [h]ow many people who test positive for the disease will actually have the disease? ____ out of ____.” *Id.* at 24-25.

⁴⁰¹ *Id.* at 25-26, 31.

⁴⁰² *Id.* at 26-28, 31 (“[A]sking these questions seemed to clarify the problem for subjects.”).

⁴⁰³ *Id.* at 31-33.

⁴⁰⁴ *Id.* at 34-37. See *id.* at 37 (“[P]resent[ing] the problem information as percents and ask[ing] for the answer as a single-event probability, elicited Bayesian [correct] performance from only 12% of subjects”; translating the problem into “frequentist terms” elicited correct reasoning from 76% of a new set of subjects; requiring subjects “to create a concrete, visual frequentist representation” pushed “performance [of a third group] to 92%”).

Cosmides and Tooby's results led them to question Kahneman's famous dictum that, "[i]n his evaluation of evidence, man is apparently not a conservative Bayesian; he is not a Bayesian at all."⁴⁰⁵ Humans may be "good intuitive statisticians after all," the authors concluded, as long as their Bayesian instincts are triggered by the expression of probabilities across many tries rather than as a "single-event probability."⁴⁰⁶ Among subjects with "a concrete *visual* representation of a population that depicts the relevant frequencies, Bayesian performance is . . . enhanced to near perfect levels."⁴⁰⁷

Recognizing that legal verdicts must be expressed as "the probability of a single event," Fenton and Neil have developed tools and visual representations to help lawyers and jurors understand the relevant concepts in terms of frequencies, while translating results into accurate single-event conclusions. They report that in the course serving as expert consultants, they have explained the same Bayesian concepts to lawyers, judges and juries using standard mathematical equations of the sort used in this article⁴⁰⁸ and "visual explanations" like those Cosmides and Tooby used in their experiments.⁴⁰⁹ "Whereas [trial participants] find it hard both to 'believe' and reconstruct formulaic explanations, they . . . understand the visual explanation."⁴¹⁰

Fenton and Neil next consider why courts perceive a difference between letting jurors use a pocket calculator to compute damages generated by the parties' competing theories while forbidding them to use Bayesian algorithms to assess the parties' competing theories about what the evidence proves.⁴¹¹ The algorithms in both tools are accepted by scientists, yet in the former case, courts don't even require expert testimony to explain the algorithms (which can generate slightly different outcomes given different rounding strategies), while in the latter case, courts forbid jurors to use the tool even after such testimony is offered. The authors conclude that the feature present in the calculator instance and missing in the case

⁴⁰⁵ Daniel Kahneman & Amos Tversky, *Subjective Probability*, supra note 184, at 450.

⁴⁰⁶ Cosmides & Tooby, supra note 393, at 59, 69.

⁴⁰⁷ *Id.* at 59 (emphasis in original).

⁴⁰⁸ See supra notes 35-42 and accompanying text.

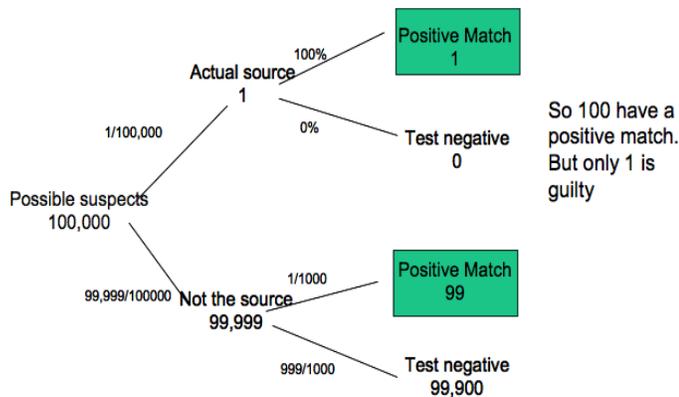
⁴⁰⁹ See Fenton & Neil, *Avoiding Fallacies*, supra note 113, at 5-6, 17-23; Fenton & Neil, *Jury Fallacy*, supra note 393, at 4, 7-11.

⁴¹⁰ See Fenton & Neil, *Avoiding Fallacies*, supra note 113, at 17.

⁴¹¹ *Id.* at 15-16.

of Bayesian analysis is the willingness of lawyers and jurors to accept, because they themselves can perform, the relevant type of calculation. To replicate that feature, Fenton and Neil developed visual tools to enable trial participants to understand and generate simple outcomes with the Bayesian algorithm. One tool is an animated version of a decision tree used in medical decision making which they use to display a simple Bayesian example in which there are 100,000 suspects and a forensic match of a rare blood type present in a only one-tenth of one percent of the population. In formulaic terms, the Bayesian equation is $1/99,999 \times 1/0.001 = 1/99.99 \rightarrow 1\%$. Fenton and Neil use the decision tree in lieu of the formula (see *Figure 3*, sans animation) to show that, as rare as the blood type is, the odds that the defendant is the culprit are still very low (1 to 99).⁴¹²

Figure 3



Building on the “chart” strategy from the paternity cases, which gives jurors a list of the Bayesian outcomes impelled by any prior odds of paternity from 1 to 100 percent,⁴¹³ Fenton and Neil give jurors an electronic Bayesian calculator (a version of which is called a Hugin⁴¹⁴) and let them use it to see the effect of different assumptions. For example, jurors might decide there are more suspects (decreasing the prior

⁴¹² Id. at 17. The authors round the odds to avoid requiring jurors to imagine fractions of people.

⁴¹³ See supra note 156, 389 and accompanying text.

⁴¹⁴ See Fenton & Neil, Jury Fallacy, supra note 393, at 13 (“[B]y using Bayesian nets and a tool such as Hugin, it is possible to show all of the implications and results of a complex Bayesian argument without requiring any understanding of the underlying . . . mathematics.”); <http://www.hugin.com/> (calling itself “the leading decision support tool”).

odds) or that there's a chance the defendant's blood got onto the victim by accident (increasing the denominator). The researchers then propose expert testimony or instructions telling jurors that, "although we were able to explain this to you from scratch, there is a standard calculation engine (accepted and validated by the entire mathematical . . . community) which will do the calculation . . . instantly for us . . . much like relying on a calculator to do long division." The jury can assume the calculator's accuracy and focus on making good assumptions about the ingredient probabilities.⁴¹⁵

Fenton and Neil imagine the application of these tools in the *Adams* case discussed above⁴¹⁶ to help a jury aggregate probabilities associated with the three items of evidence in the case: a "big" evidence match between Adams and DNA in semen on the rape victim, as to which the two sides offered conflicting expert testimony on the proper random-match frequency,⁴¹⁷ and two non-exclusionary non-matches—Adams's alibi, and the rape victim's description of the assailant as fifteen years younger than Adams whom she could not identify in a line-up.⁴¹⁸ Using a suspect pool of 200,000 based on evidence in the case, the example demonstrates the effect of the big evidence by itself (a 91% probability of guilt using the random-match frequency offered by the defense; a 99.8% probability of guilt using the frequency offered by the prosecution expert) and the effect of that evidence together with conservative estimates of the effect of the non-matches (the probability of guilt drops from 91% to 36% if the jury accepts the defense random-match frequency for the DNA, but only from 99.8 to 98.2% using the prosecution's frequency).⁴¹⁹ The jury then decides "if the assumptions in the model are reasonable" and if "the resulting probability of guilt leaves room for [reasonable] doubt."⁴²⁰

Fenton and Neil suggest limiting these tools initially to cases involving at least one item of evidence for which the random-match frequency is known or can be mined from large data bases of

⁴¹⁵ Fenton & Neil, *Avoiding Fallacies*, supra note 113, at 20.

⁴¹⁶ See supra note 43-57 and accompanying text.

⁴¹⁷ The defense and prosecution proposed random-match frequencies of, respectively 1 to 2 million and 1 to 200 million.

⁴¹⁸ Fenton & Neil, *Avoiding Fallacies*, supra note 113, at 23.

⁴¹⁹ *Id.* at 22-23.

⁴²⁰ *Id.* See also *id.* at 17-18, 23-28 (applying tools to medical malpractice case and criminal case involving blurry photo showing some but not all of the numbers on a car's license plate).

reasonably similar reference samples. Even then, of course, a given lawyer's use of the tools may be erroneous or exaggerated—failing, for example, to acknowledge that a witness could be lying or that a defendant may have been framed. But lawyers entice jurors into similar mistakes all the time, and we have no trouble relying on objections, cross-examination, counter-experts, closing argument, and other adversarial antidotes that, as we have shown, quickly curbed parallel problems with early uses of the multiplication rule in connection with inculpatory DNA.⁴²¹ Accordingly, when evidence with ascertainable random-match frequencies is available—which increasingly will be true as the legal market for data-mined frequencies expands⁴²²—it is no longer fanciful to imagine effective use of Bayesian analysis in court to reveal the aggregate value of non-exclusionary non-matches via tools no more controversial than calculators. As the next section develops, moreover, Bayesian analysis in court need play only a supporting role in effective regulatory solutions to the undervaluation of “small” non-matches.

B. Management-Based Regulation of Non-Exclusionary Non-Matches as Evidence of Identity

Using the adversarial system to accommodate aggregative analysis of non-exclusionary non-match evidence to the decision making needs of trial actors and proceedings is only a start. More systematic change is necessary to mitigate structural obstacles and legal resistance to “small” evidence of innocence and extend reforms to the crucial stages before the process becomes adversarial. Here, too, new tools are available: regulatory methods that work well when what is possible, and tailored strategies for getting there, are not yet fully understood.

1. *Management-Based Regulation*. — Professors Coglianese and Lazer distinguish three types of regulation.⁴²³ “Technology based” regulation requires regulated entities to adopt a specified solution to a given problem,⁴²⁴ such as catalytic converters for all cars or child-guards for apartment windows above a certain floor. One-size-fits-all solutions work well under relatively uniform and stable conditions but are

⁴²¹ See supra note 118-143 and accompanying text.

⁴²² See supra note 175-177 and accompanying text.

⁴²³ See Gary Coglianese & David Lazer, *Management-Based Regulation: Prescribing Private Management to Achieve Public Goals*, 37 L. & Soc'y Rev. 691, 694 (2003).

⁴²⁴ See id. at 694.

over- and under-inclusive when conditions vary and can discourage innovation by tying actors to outdated technologies.⁴²⁵ Alternatively, “performance based” regulation imposes floors or ceilings on regulated outputs.⁴²⁶ Limits on pollution factories may emit and average test scores schools’ students must attain are examples. Performance-based regulation allows actors to adopt technologies suited to varying conditions. It struggles, however, if the desired end state or effective ways to measure it are unknown and may continue to treat regulated entities as similar when they face disparate challenges, as when schools are held to the same average proficiency levels, though their students arrive at vastly different starting points.⁴²⁷

By contrast, management-based regulation induces regulated entities to develop their own strategies, which can be precisely customized to local conditions and can expose previously unknown end-state possibilities, ways to get there and measures of success.⁴²⁸ Regulators, for example, may require regulated entities to generate plans to avoid harmful results, monitor implementation and outcomes, and impose consequences to motivate actors to adjust plans in light of evidence of success or failure.⁴²⁹ Planning also can be induced by publicizing undesirable outcomes such as toxic releases,⁴³⁰ publically comparing results attained by similarly situated entities to expose less and more effective operations,⁴³¹ and using “penalty defaults” to threaten entities with severe or unpredictable consequences unless they agree to solutions with other stakeholders.⁴³² Management-based strategies work best when desired

⁴²⁵ See *id.* at 701.

⁴²⁶ See *id.*

⁴²⁷ See *id.*

⁴²⁸ See *id.* at 702-03.

⁴²⁹ See *id.* at 694, 713-14 (comparing planning requirements that give regulated entities flexibility in identifying harms to others specifying stages of the regulated activity the plan must address); *id.* at 706-11 (advocating choices among different types of planning and implementation mandates, depending on different incentives regulated entities have or can be given to plan and implement).

⁴³⁰ See, e.g., Charles Sabel et al., *Beyond Backyard Environmentalism: How Communities Are Quietly Refashioning Environmental Regulation*, *Boston Review*, Oct./Nov. 1999, <http://bostonreview.net/BR24.5/sabel.html> (discussing federal Toxic Release Inventory and Massachusetts Toxic Use Reduction Act).

⁴³¹ See, e.g., James S. Liebman & Charles Sabel, *The Federal No Child Left Behind Act and the Post-Desegregation Civil Rights Agenda*, 81 *N.C. L. Rev.* 1703, 1708-20 (2003).

⁴³² See, e.g., Bradley C. Karkkainen, *Information Forcing Regulation and Environmental Governance*, in *Law and New Governance in the E.U. and the U.S.* 301-05 (Grainne de Burca & Joanne Scott eds. 2007); Charles Sabel & William Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 *Harv. L. Rev.* 1015, 1067 (2004).

outcomes and possible solutions are uncertain or when conditions faced are highly variable.⁴³³ Both conditions characterize the criminal justice practices that invite heuristic biases and discourage use of non-exclusionary non-matches.

First, with the recent exception of DNA and other forensic evidence,⁴³⁴ the idea of systematically regulating the collection, documentation, analysis, data-mining for random-match frequencies, and presentation of clues to criminal identity is new. There is no inventory of preferred strategies or consensus on how to measure success.⁴³⁵ For example, criminal investigators know they can increase the chance of finding identifying traces by superimposing a fine-mesh grid on relevant surfaces and minutely examining and photographing each cell, but standards for when to use the technique and how well it was deployed are not widely used.⁴³⁶ Additionally, given the speed with which new technologies have developed under the impetus of adversarial testing for forensically analyzing DNA and other traces, mining burgeoning data repositories for random-match frequencies and using “frequentist” and graphical representations and calculation tools to demystify Bayesian analysis,⁴³⁷ it is too soon to mandate particular technologies or performance levels. Instead, state and local customization of solutions to distinct conditions is a more effective invitation for innovation.

2. *Regulation of “Big” Evidence of Identity.* — As Michael Dorf and Katherine Kruse have shown,⁴³⁸ management-based regulation of criminal justice activities traces back to two famed mid-1960s Supreme Court decisions regulating the most common and dangerously over-valued types of “big”

⁴³³ See Coglianese & Lazer, *supra* note 423, at 704-06.

⁴³⁴ See, e.g., Rhonda W. Wheat & Allan Jamieson, *A Tale of Two Approaches—The NAS Report and the Law Commission Consultation Paper on Forensic Science*, 7 *Int’l Comm. Evid.* (2009) (comparing The Forensic Science Regulator (UK), Review of the Options for the Accreditation of Forensic Practitioners (January 2009) and National Academy of Sciences, *Strengthening Forensic Science in the United States: A Path Forward* (2009)); sources cited *supra* note 136.

⁴³⁵ Cf. Andrew Gelman et al., *A Broken System: The Persistent Patterns of Reversals of Death Sentences in the United States*, 1 *J. Empirical Legal Stud.* 209 (2004) (using judicial reversal rates to estimate accuracy of capital verdicts).

⁴³⁶ See Daniel O. Larson et al., *Advanced Scientific Methods and Procedures in the Forensic Investigation of Clandestine Graves*, 27 *J. Contemp. Crim. Just.* 149, 153 (2011) (describing grid and excavation methods forensics teams can use to avoid compromising crime scenes).

⁴³⁷ See *supra* notes 118-178, 392-422 and accompanying text.

⁴³⁸ Michael C. Dorf & Barry Friedman, *Shared Constitutional Interpretation*, 2000 *Sup. Ct. Rev.* 61, 81-82 (2000); Michael C. Dorf & Charles F. Sabel, *A Constitution of Democratic Experimentalism*, 98 *Colum. L. Rev.* 267, 318 (1998); Kruse, *supra* note 18, at 648, 670-73 & n.11.

evidence: custodial confessions and eyewitness identifications. Although *Miranda v. Arizona* and *United States v. Wade* are best remembered for procedures that go by their names—*Miranda* warnings about a suspect’s right to counsel and to remain silent during interrogation,⁴³⁹ and *Wade* hearings enforcing a right to counsel and protection against suggestive line-ups⁴⁴⁰—neither case actually mandated that procedure.⁴⁴¹ Rather, the Court tried to use a threat to overturn convictions if the procedure was not implemented as a “penalty default” to induce jurisdictions to adopt alternative, equally or more protective solutions that were better suited to local conditions. Jurisdictions that adopted alternative measures could avoid the constitutional default rule.⁴⁴² As Professor Kruse shows, the Court issued these invitations as much out of a desire for stronger protections than was possible on a one-size-fits-all basis as out of a hope of easing burdens on law enforcement and courts.⁴⁴³

The problem with the *Miranda* and *Wade* default rules is that they were neither onerous enough to encourage local officials to “bargain around” them by adopting more tailored and efficient rules nor protective enough to impose much of a constraint on unreliably coercive interrogations and suggestive eyewitness identifications.⁴⁴⁴ Indeed, when conservative advocates made an all-out assault on *Miranda* in *Dickerson v. United States*, law-enforcement *amici curiae* were its staunchest defenders, helping convince the Court to leave *Miranda*’s default requirement in force.⁴⁴⁵ Recently, however, prompted by DNA exonerations of defendants convicted based on false confessions and inaccurate eyewitness

⁴³⁹ *Miranda v. Arizona*, 384 U.S. 436, 478-79 (1966).

⁴⁴⁰ *United States v. Wade*, 388 U.S. 218, 228 (1967).

⁴⁴¹ Kruse, *supra* note 18, at 670-73.

⁴⁴² *United States v. Wade*, 388 U.S. at 239 (inviting “[l]egislative or other regulations, such as those of local police departments, which eliminate the risks of abuse and unintentional suggestion at lineup proceedings,” which, if adopted, would “remove the basis for regarding the stage as ‘critical,’” obviating the requirement of counsel); *Miranda v. Arizona*, 384 U.S. at 467, 490 (imposing “*Miranda* warnings” requirement only on states that did not “develop their own safeguards” that “are at least as effective”).

⁴⁴³ Kruse, *supra* note 18, at 671-72 (arguing that *Miranda* and *Wade* encouraged “legislatures and law enforcement agencies to take their own steps to improve police investigatory practices” because the Justices knew the prophylactic measures they proposed were “a pale substitute for improving the police procedures themselves” and that it was “impossible for us to foresee the potential alternatives for protecting the privilege which might be devised by Congress or the States in the exercise of their creative rule-making capacities” (quoting *Miranda*, 384 U.S. at 467; citing *Wade*, 388 U.S. at 239)).

⁴⁴⁴ See, e.g., Paul G. Cassell, *Miranda*’s Social Costs: An Empirical Reassessment, 90 Nw. U. L. Rev. 387, 477-78, 486-97 (1996); Kruse, *supra* note 18, at 666-72 (discussing literature).

⁴⁴⁵ *Dickerson v. United States*, 530 U.S. 428, 444, 438 (2000).

identifications, a number of states have begun regulating interrogations and line-ups. Much of the regulation has been technology-based, requiring videotaped confessions and sequential double-blind line-ups,⁴⁴⁶ but Wisconsin and Texas have adopted more comprehensive and flexible management-based solutions.⁴⁴⁷

The Wisconsin Supreme Court's 2005 decision in *State v. Dubose* triggered that state's new approach, by interpreting the state constitution to impose stricter admissibility standards for eyewitness identifications than the federal constitutional rule.⁴⁴⁸ Under the federal rule, identifications based on suggestive procedures are admissible if there are other indicia of "reliability," such as a witness's certainty at the time of the line-up (a factor studies actually correlated with error).⁴⁴⁹ Under the Wisconsin rule, all identifications produced by "unnecessarily suggestive" procedures are inadmissible.⁴⁵⁰

To help local law enforcement agencies cope with *Dubose's* narrow and unpredictable admissibility standard, the Wisconsin Legislature required each agency to devise a written policy "to reduce the potential for erroneous identifications."⁴⁵¹ Agencies had to base their plans initially on "model policies," effective strategies from "other jurisdictions" and practices shown by research to improve "objectivity and reliability," then to revise their plans every two years based on evolving local and statewide experience.⁴⁵² For guidance, the Wisconsin Department of Justice promulgated and periodically updates a Model Policy and Procedure for Eyewitness Identification, citing studies supporting each

⁴⁴⁶ See, e.g., Ohio Rev. Code Ann. § 2933.83 (West 2010) (requiring most live and photo line-ups to be conducted by officials unaware of which participants are suspects); Gershel, *supra* note 23, at 11, 17, 19, 20, 23, 24, 35 (citing legislation and regulations); Suzanne Smalley, Police Update Evidence Gathering: Suspect Identification is Focus of Change, *Boston Globe*, July 20, 2004, at B1 (listing jurisdictions requiring double-blind line-ups); sources cited *supra* notes 21-23.

⁴⁴⁷ H.B. 215, 82nd Leg., Reg. Sess. (Tex. 2011) (signed by Gov. Rick Perry, June 17, 2011) (requiring law enforcement agencies to adopt standards and procedures for live and photographic line-ups designed by the Law Enforcement Management Institute of Texas or implement their own equally or more protective policy); *infra* notes 451-456 and accompanying text (discussing Wisconsin).

⁴⁴⁸ 699 N.W.2d 582, 596-97 (Wis. 2005).

⁴⁴⁹ See *Perry v. New Hampshire*, 131 S. Ct. 2932, ____ (2012); *Manson v. Braithwaite*, 432 U.S. 98, 106, 114 (1977); *Neil v. Biggers*, 409 U.S. 188, 199 (1972). Cf. *Kruse*, *supra* note 18, at 669-70 (criticizing "reliability" escape hatch for using factors research associates with inaccuracy as a basis for admitting suggestive identifications).

⁴⁵⁰ *State v. Dubose*, 699 N.W.2d at 592-93.

⁴⁵¹ Wis. Stat. § 175.50(2).

⁴⁵² *Id.* §§ 175.50(2)-(5) (2005-2006) (inviting but not requiring agencies to adopt double-blind and sequential procedures).

component.⁴⁵³ As Professor Kruse points out, the Wisconsin reforms use *Dubose*'s novel standard and exclusionary rule as a penalty default to encourage law enforcement agencies to create safe harbors by adhering to locally devised plans that are continuously benchmarked against a state model, the best available research, and proven practices from elsewhere in the state and nation.⁴⁵⁴

Wisconsin took a similar path to regulate confessions. Reacting to disturbing tactics used by Milwaukee police to induce a fourteen-year-old to confess, the Wisconsin high court found a constitutional violation under the federal "involuntariness" standard, then ruled that, henceforth, juvenile confessions would be inadmissible under state law unless they were videotaped or taping was infeasible.⁴⁵⁵ Expecting the court to extend similarly inflexible, technology-based regulation to adult confessions, the Wisconsin Legislature codified the rule requiring taping of custodial confessions by juveniles and adopted a new "policy" encouraging videotaping of adult confessions by requiring jurors to be instructed to "consider the absence of recording" when evaluating custodial statements.⁴⁵⁶ The latter procedure puts police agencies to a management-based choice: either videotape interrogations or adopt alternative procedures that jurors believe are equally productive of reliable confessions.

Wisconsin's legislation governing identifications and confessions aims to motivate law enforcement agencies to disclose information about their practices, which the Wisconsin Department of Justice can compare to practices of and share with other agencies statewide.⁴⁵⁷ As Professor Kruse notes, such regimes face a tension between goals of transparency (allowing inter-jurisdictional comparison and

⁴⁵³ See, e.g., Wisconsin Dep't of Just. Bur. of Training and Standards for Crim. Just., Model Policy and Procedure for Eyewitness Identification 2 (Apr. 1, 2010), <http://www.doj.state.wi.us/dles/tns/eyewitnesspublic.pdf> ("[E]yewitness evidence is much like trace evidence left at a crime scene [that can be contaminated] if not handled properly" (citing Wells & Olson, *supra* note 18, at 286-89)); Wisconsin Dep't of Justice Bur. of Training & Standards for Crim. Just., Comprehensive Review and Analysis of Best Practices (2010), www.doj.state.wi.us/dles/tns/eyewitnesspublic.pdf (urging investigators to avoid "fishing expeditions" by delaying lineups until alibis and competing leads are investigated and forensic testing is complete); Kruse, *supra* note 18, at 648 n.10, 687.

⁴⁵⁴ Kruse, *supra* note 18, at 685, 689-90.

⁴⁵⁵ *In re Jerrell C.J.*, 699 N.W.2d 110, 113 (Wis. 2005).

⁴⁵⁶ Wis. Stat. § 938.31(3)(b)-(c) (codifying *Jerrell* rule); *id.* § 968.073(2) (establishing "policy" in felony cases of "audio or audio and visual recording of a custodial interrogation"); *id.* § 972.115(2)(a) (specifying jury instruction when unrecorded custodial statements are admitted); see Kruse, *supra* note 18, at 690-93.

⁴⁵⁷ See Kruse, *supra* note 18, at 727-28 ("This new body of information provides rich opportunities for Wisconsin to pursue . . . continuous improvement, public accountability, and cross-jurisdictional learning.").

sharing) and accountability (creating disincentives to collect and publicize information that may be used against the agency in court).⁴⁵⁸ Wisconsin's solution to this tension, characteristic of management-based strategies, is a penalty default: an effective but onerous and inflexible procedural burden that induces agencies to devise alternative strategies and disclose their results in an effort to avoid the default rule.⁴⁵⁹

3. *Regulation of "Small" Non-Matches* —Analogous management-based strategies can also induce criminal process actors to use "small," non-exclusionary non-match evidence to improve the accuracy of identity determinations, including planning mandates, statewide benchmarks and publication of comparative error rates at each stage of the investigative and trial process. For example, agencies could be put to the choice of adopting other localities' best practices or justifying their own actions to juries by allowing the defense to admit evidence of policies in use by comparable law-enforcement agencies in the state for handling non-exclusionary non-matches and to instruct juries to consider an agency's failure to use effective practices in evaluating the evidence.⁴⁶⁰ Numerous improvements are available to motivated agencies:

- Comprehensive videotaping of crime scenes, with technology for acute magnification and other aids to after-the-fact review and preservation of evidence and monitoring of investigative rigor⁴⁶¹
- Institutional structures or double-blind procedures to isolate pretrial investigative, forensic-analysis and accusatory functions and temper "uniqueness," "irrelevance" and other heuristic mistakes triggered by early identification of scenarios and suspects⁴⁶²

⁴⁵⁸ See id. at 728, 731.

⁴⁵⁹ See, e.g., Sabel et al., *supra* note 430, at 5-6 (describing regulatory power of public disclosure of comparative toxic-release data); Joseph Rees, *Hostages of Each Other: The Transformation of Nuclear Power Safety After Three Mile Island* 118-20 (1994) (describing effect of annual disclosure of comparative error-rate data at meeting of all CEOs of nuclear power companies, even absent public disclosure of the data).

⁴⁶⁰ See *supra* notes 451-459 and accompanying text.

⁴⁶¹ See, e.g., Edward Robinson, *Crime Scene Photography* 582 (2007) (describing value of photography for crime-scene preservation); Dan L Warnock, *The Iraqi Criminal Justice System, an Introduction*, 9 *Denv. J. Int'l L. & Pol'y* 1, 24 (2010) (listing practices U.S. urged Iraqi investigators to adopt to document crime scenes with photographs, diagrams and tape recordings); Dick Warrington, *Crime Scene Photography: Capturing the Scene*, *Forensic Magazine*, August/September 2009, <http://www.forensicmag.com/article/crime-scene-photography-capturing-scene> (similar).

⁴⁶² See, e.g., Findley & Scott, *supra* note 208, at 393-94 (discussing value of separating investigative and charging functions to avoid heuristic bias).

- Devil’s advocate mechanisms, prompting police before turning cases over to prosecutors, and prosecutors before pressing charges to inventory alternative scenarios about who committed the crime and innocent explanations for evidence that does and does not match suspects⁴⁶³
- Training and expert assistance for police, forensic analysts, prosecutors and public defenders in graphics- and calculator-aided use of Bayesian analysis of the conjoint inculpatory and exculpatory power of evidence for which frequency information is available and in estimating and comparing the effect of different prior odds⁴⁶⁴
- Training for the same actors in data-mining techniques for generating frequency information about potentially matching and non-matching clues⁴⁶⁵
- Open-files discovery of all non-exculpatory non-matches;⁴⁶⁶ enforcement of ABA Model Rule 3.8(d) (which most states follow but do not apply to prosecutors) obliging state disclosure of all exculpatory evidence⁴⁶⁷
- Admissibility at trial, unless prejudice substantially outweighs probative value, of evidence of sins of commission and omission in the state’s capture, documentation, forensic analysis and presentation of trace evidence and witness information that could identify the perpetrator, and instructions allowing juries to treat either failing as sufficient to establish a reasonable doubt as to guilt⁴⁶⁸

⁴⁶³ See, e.g., Molly Johnson & Laural Hooper, Resource Guide for Managing Capital Cases: Federal Death Penalty Trials 10-11 (Fed. Judicial Center 2004) (describing multi-tiered process for approving federal death-penalty prosecutions).

⁴⁶⁴ See supra notes 392-422 and accompanying text.

⁴⁶⁵ See supra notes 175-177 and accompanying text.

⁴⁶⁶ See, e.g., Avis E. Buchanan, Fairer Trials and Better Justice in D.C., Wash. Post, Oct. 28, 2011, http://www.washingtonpost.com/opinions/fairer-trials-and-better-justice-in-dc/2011/10/25/gIQA__FMQM_story.html (advocating open-files discovery in use in several jurisdictions).

⁴⁶⁷ See ABA Comm. on Ethics and Prof ’l Resp., Formal Op. 09-454, at 1, 4-5 (2009) (interpreting Model Rule of Professional Conduct Rule 3.8(d), requiring “timely disclosure to the defense of all evidence or information known to the prosecutor that tends to negate the guilt of the accused,” as broader than the *Brady* rule because it applies irrespective of “the anticipated impact of the evidence” on trial outcomes); Kevin C. McMunigal, The (Lack of) Enforcement of Prosecutorial Rules, 38 Hofstra L. Rev. 847, 850-55, 860-64 (2010) (discussing states’ failure to enforce ABA Opinion 9-454). See also Attorney General (U.K.), Attorney General’s Guidelines on Disclosure 2 (2005), <http://www.attorneygeneral.gov.uk/Publications/Documents/disclosure.doc.pdf> (last visited June 7, 2011) (requiring disclosure of “all material held by the prosecution that weakens its case or strengthens that of the defence”).

⁴⁶⁸ See supra notes 316-334 and accompanying text.

- Admissibility, subject to the same restriction, of evidence that an alternative suspect committed the crime, including non-exclusionary non-matches as to the defendant that match the alternative suspect, and jury instructions to consider alternative scenarios and hypotheses suggested by that evidence⁴⁶⁹
- Admissibility at trial of expert testimony facilitating understanding and use of, and calculator tools for implementing, Bayesian analysis⁴⁷⁰
- Broad admissibility of data-mined random-match frequencies, subject to broad adversarial testing.⁴⁷¹

The goal here is not to privilege any particular step and instead to show the wealth of planning regimes, penalty defaults, monitoring mechanisms, enforcement techniques and improvement steps that courts, legislatures and agencies can adopt to improve the use of non-exclusionary non-match and other small (and big) evidence and boost the accuracy of criminal identity determinations. If there is a will, there are many ways.

VII. Conclusion

Inattentiveness to small flecks of non-matching evidence is no less implicated in the miscarriages of justice exposed by recent DNA exonerations than is excessive attention to the fool's gold of suggestive eyewitness identifications, pressured confessions and misinterpreted forensic evidence. Indeed, the two problems are opposite sides of the same trick coin, which dupes intuitive decision makers, via "uniqueness" and "irrelevance" manifestations of the "certainty" fallacy and other heuristic mistakes, into falsely treating eyewitness identifications, confessions, fingerprints and other "big" evidence as matching a unique trait of the perpetrator and the defendant and treating "small" evidence of non-matching traces as so easily explained away that they are irrelevant. In fact, all identity evidence—eyewitness testimony and confessions, no less than fingerprints and DNA, and all manner of so-called "circumstantial" minutia—

⁴⁶⁹ See supra notes 353-377 and accompanying text.

⁴⁷⁰ See supra notes 378-422 and accompanying text.

⁴⁷¹ See supra notes 122-143 and accompanying text.

acquires its strength through the aggregation of individually unimpressive probabilities associated with matches or non-matches between clues and suspects.

Motivating investigators, forensic analysts, prosecutors, defense lawyers, judges and jurors to give appropriately disciplined attention to aggregations of the many small probabilities that should guide decisions on the way from crime to punishment is difficult but not impossible, individually or institutionally. Aided by a little patience and a panoply of old-fashioned adversarial methods of bringing to mind competing, uniqueness- and irrelevance-disproving stories, a multitude of new-fangled strategies for mining ubiquitous data for frequency information, making Bayesian analysis accessible to all audiences, and inducing institutions to ever-more-effectively manage risk, we have all the tools we need to fashion minute probabilistic flecks into treasurably reliable criminal verdicts.