2012

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

James S. Liebman
*Columbia Law School*, jliebman@law.columbia.edu

Shawn Blackburn
author1952729@ssrn.com

David Mattern
author1952730@ssrn.com

Jonathan Waisnor
author1952731@ssrn.com



Recommended Citation

James S. Liebman, Shawn Blackburn, David Mattern & Jonathan Waisnor, *The Evidence of Things Not Seen: Non-Matches as Evidence of Innocence*, Iowa Law Review, Vol. 98, p. 577, 2012; Columbia Public Law Research Paper No. 13-333 (2012).
Available at: https://scholarship.law.columbia.edu/faculty_scholarship/1776

# Columbia Law School

Public Law & Legal Theory Working Paper Group

Paper Number 13-333

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

James S. Liebman
Columbia Law School

and

Shawn Blackburn
J.D., Columbia Law School, 2011

David Mattern
J.D. Candidate, Columbia Law School, 2013

Jonathan Waisnor
J.D., Columbia Law School, 2012

February 20, 2013

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

*James S. Liebman,* Shawn Blackburn,** David Mattern*** & Jonathan Waisnor*****

*ABSTRACT: 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 can be telling evidence of innocence (e.g., an aspect of an eyewitness's description of the perpetrator that does not match the suspect she identifies in a lineup, or a loose button found at the crime scene that does not match the suspect's clothes). Although non-matching clues often are easily explained away, making them seem uninteresting, they frequently turn out to match the real culprit when exonerations reveal that the wrong person was convicted. This Article uses "non-exclusionary non-matches" and what would seem to be their 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 those forms of evidence seem stronger than they are, while, conversely, (3) our minds and procedures magnify the circumstantial character of non-exclusionary non-matches, making them seem weaker than they are. We propose ways to use circumstantial matches and non-matches more effectively to avoid miscarriages of justice.*

† *See* JAMES BALDWIN, THE EVIDENCE OF THINGS NOT SEEN (1985).

* Simon H. Rifkind Professor, Columbia Law School. Thanks to Alexandra Blaszczuk, Leslie Demers, Robert King, and Abshir Kore for superb research assistance. This Article is dedicated to David Baldus, who until his recent death was the Joseph B. Tye Professor of Law at The University of Iowa. As a pioneer in bringing the discipline of science and rigorous inquiry—and the courage to confront the facts as best we can discern them and as troubling as they may be—to the analysis of the operation of the criminal justice system, Professor Baldus was an inspiration for this Article and for all of our work. We are deeply saddened by his death.

** J.D., Columbia Law School, 2011.

*** J.D. Candidate, Columbia Law School, 2013.

**** J.D., Columbia Law School, 2012.

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

In 1998 in *State v. Hayes*, a New Orleans jury convicted juveniles Ryan Matthews and Travis Hayes of the robbery–murder of a New Orleans store owner.[1] 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.[2] Police stopped Hayes and Matthews in a vehicle resembling witnesses' description of the getaway car.[3] Hayes confessed to the police that he drove the getaway car, and that Matthews, who was high on marijuana, entered the store and ran out after several gunshots.[4] A witness identified Matthews as the man he saw running from the store.[5]

Although strong, the evidence was deficient in some respects. Hayes was borderline mentally retarded and confessed after six hours of interrogation.[6] His statements were modestly inconsistent with each other and with known details of the crime.[7] The make and model of the car Hayes was driving when 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.[8] The eyewitness who identified Matthews watched the events through a rear-view mirror,[9] and other eyewitnesses could not identify Matthews and described the shooter as 5′6″— six inches shorter than Matthews.[10] DNA on a ski mask left at the crime scene matched neither defendant.[11] None of the deficiencies excluded the possibility of guilt, however, and the jury convicted both men. Hayes was sentenced to life in prison and Matthews to death.[12]

Shortly after the trial, Rondell Love was convicted of an unrelated manslaughter and sentenced to twenty years in the Louisiana State Penitentiary, where Matthews was awaiting execution.[13] Love told other

1. *See* State v. Hayes, 2001-736, p. 2–3 & n.1 (La. App. 5 Cir. 12/26/01); 806 So. 2d 816, 818 & n.1; Bob Herbert, Op-Ed., *Trapped in the System*, N.Y. TIMES (July 14, 2003), http://www.nytimes.com/2003/07/14/opinion/trapped-in-the-system.html; Michael Perlstein, "*A Forgotten Man": Prosecutors Refuse To Reconsider Inmate's Case Despite Evidence Supporting His Claim*, TIMES-PICAYUNE (New Orleans, La.) (Oct. 10, 2004), http://www.truthinjustice.org/travis-hayes.htm.

2. Perlstein, *supra* note 1.

3. *Id.*

4. *Hayes*, 806 So. 2d at 820–21 & nn.9–10.

5. *Id.* at 820.

6. *See* Perlstein, *supra* note 1.

7. *See id.*

8. *See id.*

9. *Hayes*, 806 So. 2d at 819.

10. Herbert, *supra* note 1.

11. Perlstein, *supra* note 1.

12. *See id.*

13. *See* Herbert, *supra* note 1.

inmates that he had committed the murder for which Matthews was condemned.[14] Matthews 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.[15] Matthews and Hayes were eventually exonerated.[16]

In 1983 in *State v. DeLuna*, a jury convicted Carlos DeLuna of stabbing a young female Hispanic clerk to death at a Diamond-Shamrock gas station in Corpus Christi, Texas.[17] 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 seen the assailant escape on foot.[18] Again, there were discrepancies, including the eyewitness's initial description of the shabbily dressed, mustachioed, and bewhiskered "derelict" he saw struggling with the victim;[19] fingerprints found 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.[20] The crime was captured on a 911 phone call from the store clerk.[21] Police arrested DeLuna three blocks away, cowering under a pick-up truck.[22] DeLuna testified that he had seen an acquaintance named 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.[23] A police officer testified that he scoured police records for a "Carlos Hernandez" matching the eyewitness description of the assailant but found none.[24] The prosecutor dubbed Hernandez a "phantom" and DeLuna

14. Perlstein, *supra* note 1.

15. *See* Herbert, *supra* note 1.

16. Elisabeth Salemme, *Gallery of the Exonerated: Travis Hayes and Ryan Matthews*, TIME, http://www.time.com/time/specials/2007/article/0,28804,1627368_1627366_1627389,00.html (last visited Oct. 3, 2012) (noting that Matthews was exonerated in June 2004 and released in December 2006, following eight years in prison).

17. *See* James S. Liebman et al., *Los Tocayos Carlos*, 43 COLUM. HUM. RTS. L. REV. 711, 724, 1019 (2012); Steve Mills & Maurice Possley, "*I Didn't Do It. But I Know Who Did*," CHI. TRIB. (June 25, 2006) [hereinafter Mills & Possley, *"I Didn't Do It"*], http://www.chicagotribune.com/news/chi-tx-1-story,0,653915.story; Steve Mills & Maurice Possley, *A Phantom, or the Killer?*, CHI. TRIB. (June 26, 2006) [hereinafter Mills & Possley, *A Phantom*], http://articles.chicagotribune.com/2006-06-26/news/0606260189_1_jurors-hernandez-home-gas-station; Maurice Possley & Steve Mills, *Did One Man Die for Another Man's Crime?*, CHI. TRIB. (June 27, 2006), http://articles.chicagotribune.com/2006-06-27/news/0606270137_1_gas-station-stabbed-killing; *see also* Karen Boudrie Greig, *Last Call from Death Row: Seeking the Truth During a Final Conversation*, NEW ORLEANS MAG., Aug. 2012, at 94.

18. Mills & Possley, *"I Didn't Do It," supra* note 17.

19. *See* Liebman et al., *supra* note 17, at 909.

20. *See id.* at 912.

21. Mills & Possley, "*I Didn't Do It," supra* note 17.

22. *Id.*

23. Liebman et al., *supra* note 17, at 994.

24. *Id.* at 999.

a "liar" for fabricating him.[25] During DeLuna's appeals, an affirming court expressed "substantial doubt that Carlos Hernandez even existed."[26] DeLuna was quietly executed in 1989.[27]

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.[28] Three years before the convenience store stabbing, the detective and assistant district attorney handling that case had considered Hernandez a prime suspect in the beating and knifing death of another young Hispanic woman.[29] Although the detective and prosecutor eventually arrested and tried a different man, he was acquitted after the man's attorney marshaled evidence that 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 the gas station killing.[30]

Shortly after the killing, and for years afterwards, Hernandez told associates that he, not DeLuna, committed the crime.[31] 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.[32] Hernandez had 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.[33]

Hernandez's characteristic modes of dress and grooming at the time of the convenience store murder, and many other traits that did not match DeLuna, did match the eyewitness's initial description of the assailant.[34]

25. *Id.* at 1002.

26. DeLuna v. McCotter, No. C-86-234, slip op. at 8 (S.D. Tex. June 13, 1988), *available at* http://www3.law.columbia.edu/hrlr/ltc/sources/deluna-v-mccotter-s-d-tex-june-13-1988.pdf.

27. Mills & Possley, *A Phantom*, *supra* note 17.

28. Liebman et al., *supra* note 17, *passim.*

29. *Id.* at 720–21, 846–48, 853–66.

30. *Id.* at 898–900. The profile view in the photographs is the same one the eyewitness had as he watched the assailant flee. *See infra* Figure 1.

31. Liebman et al., *supra* note 17, at 876–82, 889–93.

32. *Id.* at 1093–98.

33. *Id.* at 1092–97.

34. *See infra* Figure 2 (analyzing matching and non-matching evidence in *DeLuna*). Hernandez died in prison in 1999. Liebman et al., *supra* note 17, at 721, 1098–99.

FIGURE 1

Carlos DeLuna (left) two weeks before the killing for which he was executed. Carlos Hernandez (right) two months after that killing, upon his arrest 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 *Matthews/Hayes* and *DeLuna*, commentators typically 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's confession and the show-up identification of DeLuna.[35] These observers point out that identifications,[36] confessions,[37] and "snitch"

35. *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].'" (quoting Bruton v. United States, 391 U.S. 123, 139–40 (1968) (White, J., dissenting))); Colorado v. Connelly, 479 U.S. 157, 182 (1986) ("[A] confession makes the other aspects of a trial in court superfluous, and the real trial, for all practical purposes, occurs when the confession is obtained." (quoting MCCORMICK ON EVIDENCE 316 (Edward W. Cleary ed., 2d ed. 1972)) (internal quotation marks omitted)).

36. *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 approximately 35% of 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); 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 mistaken-identity literature); Sandra Guerra Thompson, *Beyond a Reasonable Doubt? Reconsidering Uncorroborated Eyewitness Identification Testimony*, 41 U.C. DAVIS L. REV. 1487 (2008); Gary L. Wells & Elizabeth A. Olson, *The Other-Race Effect in Eyewitness Identifications: What Do We Do About It?*, 7 PSYCHOL. PUB. POL'Y & L. 230 (2001).

37. *See, e.g.*, GARRETT, *supra* note 36, at 22–33 (describing interrogation procedures eliciting false confessions); Heller, *supra* note 36, at 248–49, 254 (summarizing research

testimony[38] (another classic example 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,[39] cautionary instructions about informant testimony,[40] or exclusion of evidence unless it was collected through state-of-the-art techniques such as double-blind lineups and videotaped confessions.[41]

Exoneration cases such as *Hayes/Matthews* and reinvestigation cases such as *DeLuna* reveal a second type of indicative evidence, which commentators typically ignore. In almost all of 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 a 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

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.

38. *See, e.g.*, GARRETT, *supra* note 36, 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 Yaroshefsky, *Cooperation with Federal Prosecutors: Experiences of Truth Telling and Embellishment*, 68 FORDHAM L. REV. 917, 953 (1999).

39. *See, e.g.*, James M. Doyle, *Learning from Error in American Criminal Justice*, 100 J. CRIM. L. & CRIMINOLOGY 109 (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, 744–47 (1993) (advocating use of expert testimony on difficult memorization process required in making identifications and effect of light, movement, duration of encounter, and weapons); *cf.* People v. Santiago, 958 N.E.2d 874, 880–84 (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).

40. *See, e.g.*, Aaron M. Clemens, Note, *Removing the Market for Lying Snitches: Reforms To Prevent Unjust Convictions*, 23 QUINNIPIAC L. REV. 151, 235–42 (2004) (proposing polygraph examinations and other ways to deter false statements by jailhouse informants).

41. *See, e.g.*, Alan M. Gershel, *A Review of the Law in Jurisdictions Requiring Electronic Recording of Custodial Interrogations*, 16 RICH. J.L. & TECH. 1 (2010) (advocating videotaping of interrogations); Kruse, *supra* note 36, at 660 & n.71, 661–62 (discussing videotaping of interrogations and sequential, double-blind eyewitness lineups); Thomas P. Sullivan, *Police Experiences with Recording Custodial Interrogations*, 88 JUDICATURE 132, 132–33 (2004); John Schwartz, *Changes to Police Lineup Procedures Cut Eyewitness Mistakes, Study Says*, N.Y. TIMES (Sept. 18, 2011), http://www.nytimes.com/2011/09/19/us/changes-to-police-lineup-procedures-cut-eyewitness-mistakes-study-says.html (describing evidence that sequential, double-blind lineups "catch fewer innocent suspects"); John Schwartz, *New Orleans Police, Mired in Scandal, Accept Plan for Overhaul*, N.Y. TIMES (July 24, 2012), http://www.nytimes.com/2012/07/25/us/plan-to-reform-new-orleans-police-department.html (describing consent decree adopted by the U.S. Department of Justice and New Orleans Police Department in which the latter agreed to videotape all sexual assault and homicide interrogations from start to finish and to revise eyewitness identification procedures to avoid police influence on witnesses).

reasons having nothing to do with the crime. The police, prosecutors, and jurors in these cases 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.[42]

Examples include the non-match in the *Hayes/Matthews* 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 Matthews 6′1″ frame.[43] Examples from the *DeLuna* case are collected in Figure 2. Of twenty-nine traces associated with the crime, three match neither DeLuna (executed for the crime) nor Hernandez (who told associates he committed the crime). Seven 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 and a wad of chewed gum spat onto the floor. Only seven of the twenty-nine 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.

42. In contrast, *exculsionary* 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 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.

43. *See* Lisa L. Smith et al., *Understanding Juror Perceptions of Forensic Evidence: Investigating the Impact of Case Context on Perceptions for Forensic Evidence Strength*, 56 J. FORENSIC SCI. 409, 409 (2011) ("Current forensic science techniques are capable of recovering and 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. . . . They have decreased the quantity of trace material required to conduct useful comparative analyses.").

FIGURE 2[44]

| | **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 seen nearby** | **Total** |
|---|---|---|---|---|---|
| **Matches only DeLuna** | | | | | 0 |
| **Matches DeLuna and Hernandez** | 1. 5′8″<br>2. 160 lbs<br>3. Hispanic<br>4. Male<br>5. Black hair<br>6. Wavy hair | 14. Winston Cigarettes | | | 7 |
| **Matches only Hernandez** | 7. Mid/late 20s<br>8. Moustache<br>9. Unshaven<br>10. "Derelict" (shabby, unkempt clothes)<br>11. Red plaid flannel shirt or jacket<br>12. White or light grey sweatshirt<br>13. Blue jeans | 15. Lock-blade buck knife<br>16. Maroon button | | 29. 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 pools, smears, spatter<br>19. Cigarette fragment | 24. Bloody palm prints<br>25. Bloody shoe prints<br>26. Cement chunks<br>27. Wad of chewed gum<br>28. Beer cans (saliva) | | 7 |
| **Does not match either** | | 20. Fingerprint—phone<br>21. Fingerprint—door<br>22. Fingerprint—beer can | | | 3 |

44. The information in Figure 2 is taken from Liebman et al., *supra* note 17, at 731–84, 908–27 and figs.3, 8–10, 23 & 25–29.

It is not surprising that police, prosecutors, and jurors in the *Hayes/Matthews* and *DeLuna* cases dismissed these "small" non-matches—"small" because 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 large single vein is hard to pass up. This Article argues, however, that systematically 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 in resolving the identity question. The classic case is *People v. Collins*, 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.[45] Although decided over forty years ago and easily distinguished based on the prosecutor's flawed methodology and barely concealed racial ulterior motive, the case continues today to provide part of the intellectual basis for resisting probabilistic proof generally.

Part IV responds to the *Collins* critique by showing that the courts themselves have rejected the critique of probabilistic evidence 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 probabilistic[46] 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 being equal, the more non-unique matching

45. People v. Collins, 438 P.2d 33, 41–42 (Cal. 1968).

46. *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* note 104 and accompanying text.

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 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*'s 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*'s 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/Matthews* 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.[47]

To show how, we use a hypothetical example based on a familiar case, *People v. Adamson*.[48] 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 police found the bottom half of a pair of women's stockings nearby. The top halves of the stockings were missing. Suspicion came to rest on Dewey Adamson, a black middle-aged man. When the 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.[49]

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.[50] 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 tops. The all-white jury in the case[51] also may have seen only that type of "match" between the culprit and Adamson. The problem, of course, is that at the time, African-American men frequently used stocking tops in the process of "conking," or using chemicals to straighten their hair.[52]

---

47. *See* 1 DAVID L. FAIGMAN ET AL., MODERN SCIENTIFIC EVIDENCE: THE LAW AND SCIENCE OF EXPERT TESTIMONY app. 6A at 398–400 (2010–2011 ed. 2010) (describing deductive nature of Bayesian logic); ATHANASIOS PAPOULIS, PROBABILITY, RANDOM VARIABLES, AND STOCHASTIC PROCESSES 38–39, 78–81, 112–14 (2d ed. 1984) (deriving Bayes' Theorem mathematically).

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

49. *See Adamson II*, 332 U.S. at 59 (majority opinion) (concluding that the California court's 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 *Adamson I*, 165 P.2d at 7)).

50. *See, e.g.*, ERIC D. GREEN ET AL., PROBLEMS, CASES, AND MATERIALS ON EVIDENCE 69 (3d ed. 2000); JACK B. WEINSTEIN ET AL., CASES AND MATERIALS: EVIDENCE 6 (9th ed. 1997).

51. People v. Adamson (*Adamson III*), 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).

52. *See* AYANA D. BYRD & LORI L. THARPS, HAIR STORY: UNTANGLING THE ROOTS OF BLACK HAIR IN AMERICA 43–57 (2001); MALCOLM X WITH ALEX HALEY, THE AUTOBIOGRAPHY OF MALCOLM X 52–56 (1964).

The case also illustrates the mathematical definitions of relevance, probative weight, and prejudice that Richard Lempert famously introduced to legal audiences.[53] 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.[54] Stated mathematically, evidence is irrelevant if $P (E/G) / P (E/not\ G) = 1$. Probative weight then is the numerator divided by the denominator, except when that equals 1. 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.[55] For example, the jury may miscalculate the denominator value—the probability that the evidence would exist though the defendant is not guilty—because the jurors don't know that African-American men often use women's stocking tops as a hair-care implement.[56]

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 D. Green, Charles R. Nesson, and Peter L. Murray. Although they did not describe it quite this way, they used *Adamson* to identify an evidentiary mistake that undermines the accuracy of intuitive estimates of the numerator probability in the likelihood-ratio calculation of relevance and probative weight.[57] Building on 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:

53. Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021 (1977). In the discussion in text here, we use the conventional definition of probative weight as the likelihood ratio itself—its numerator divided by its denominator. *See, e.g.*, D.H. Kaye & Jonathan J. Koehler, *The Misquantification of Probative Value*, 27 LAW & HUM. BEHAV. 645, 649 (2003) (noting and citing authority for the proposition that "[t]he [likelihood ratio itself] is a common measure of probative value in law"). Other definitions have been proposed. *See, e.g.*, *id.* at 649 n.8 (noting statistical reasons for preferring the log-likelihood ratio as a definition of probative weight). We use a different one below, namely, the absolute value of the likelihood-ratio numerator minus its denominator divided by its numerator. *See infra* note 260 and accompanying text.

54. *See* Lempert, *supra* note 53, at 1026.

55. *See id.* at 1027.

56. Magnifying the potential for prejudice is the concern that, when the numerator value is high—when, as here, there is a high 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* Part V.A.4 (discussing the "uniqueness fallacy").

57. GREEN ET AL., *supra* note 50, at 70–71.

(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 briefly 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 twenty-nine 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, that she 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, after initially surmising that the fatal blows to the victim were from the assailant's left hand, she later backed off of that conclusion because 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 being 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, Nesson, and Murray suggested, the prosecutor's theory may entice the jury into multiple missteps.[58] 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. As Professors Green, Nesson, and Murray pointed out,[59] however, 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 of the victim, who was 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 example 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% 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% and

58. *Id.* (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 when 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) "leav[ing] 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").

59. *See id.*

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%.

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% of adult males in the area are left-handed or ambidextrous; 5% own brown sweatshirts; and 10% smoke cigarettes. Based on this data, and on the 98% 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 multiplied by 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/not G))*[60]:

*(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/not G)) = subsequent odds of guilt*

The expert treats the prior odds of guilt as 98-to-2, corresponding to the 98% probability of guilt that the jury found after hearing the state's case. The expert then analyzes the effect of each non-match, starting with the 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%. The expert then estimates the denominator probability—that the evidence would be present if someone other than Adamson committed the crime—as 15.45%. This number represents the rate of non-right-handers in the population (11%), plus the rate of right-handers (89%) multiplied by the 5% possibility that a right-hander used his left hand to beat the victim *(.05 x .89 = 4.45%)*.

60. For additional discussion of the likelihood ratio as a measure of probative value, see *supra* notes 56–59 and accompanying text.

As for the male witness's description of a short, balding forty-year-old outside the victim's home, the expert conservatively posits a 75% 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 twenty-nine 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, forty-ish" description would have been made if someone besides Adamson were guilty—slightly higher at 80%.[61] The expert likewise rates the probability of a 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% and the denominator as 20%.

Next, given that neither the victim nor Adamson smokes, the expert rates the numerator probability of finding a cigarette in the victim's house if Adamson is guilty as fairly low—say 20%. But if Adamson is not guilty, the probability is higher that the killer left a cigarette behind at the scene—say 30%. Finally, reasoning that the killer likely discarded his tell-tale outer clothing while escaping, the expert estimates a 70% 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% of men in the area wear brown sweatshirts, the denominator probability—that the police would find no brown sweatshirt on or around an innocent Adamson—is 95%.

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 reduces the odds of guilt from the jury's 98-to-2 (98%) to 15.8-to-1, or 94%:

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

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%:

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

61. If there is a 75% 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 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.

In like fashion, the shopper's encounter with a green-eyed sprinter reduces the probability to 88.1%; the cigarette reduces it to 83.2%; and the sweatshirt evidence reduces it further to 78.5%.[62]

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% 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, in part 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—especially if the jury never hears about some of them and downgrades the credibility of others due to excessive confidence in 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 *REVISITED*

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 *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.[63]

---

62. 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 |
| + 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 |

63. *See, e.g.*, MICHAEL LYNCH ET AL., TRUTH MACHINE: THE CONTENTIOUS HISTORY OF DNA FINGERPRINTING 218–19 (2008); Peter Donnelly, *Appealing Statistics*, 47 MED. SCI. & L. 14

The rape victim in *Adams* described her assailant as a white, clean-shaven twenty to twenty-five year old, with a local accent.[64] After the police arrested thirty-seven-year-old Denis Adams for reasons that are not entirely clear from the opinions in the case, the victim could not pick him out of a lineup and said he definitely was not her attacker, who was much younger.[65] The prosecution nonetheless conducted a DNA analysis, which generated a match of Adams's 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.[66] 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 markers—1 in 220—were close to 100,000 times greater than the prosecution's 1 in 200 million figure.[67] Defense experts testified that the "1 in 200 million" figure was itself 100 times too low, due to sampling errors and use of sample sizes too small to provide confidence in prosecution's estimate.[68]

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's girlfriend. Defense experts offered two responses. First, they countered with their own estimates of 1 in 2 million (assuming sampling problems were corrected) and 1 in 220 (given the possibility that Adams's brother was the culprit).[69] 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% (derived from the 1-in-220 estimate) to 1-to-2, or only 33%.[70] 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 non-

(2007); L.A. Foreman et al., *Interpreting DNA Evidence: A Review*, 71 INT'L STAT. REV. 473 (2003); J.P. Henderson, *The Use of DNA Statistics in Criminal Trials*, 128 FORENSIC SCI. INT'L 183 (2002).

64. *See* R v. Adams (*Adams I*), [1996] 2 Crim. App. 467 at 468 (Eng.).

65. *See* Donnelly, *supra* note 63, at 14.

66. *See Adams I*, 2 Crim. App. at 468 ("The prosecution case rested entirely upon expert evidence in relation to the DNA sample which was challenged by the defence.").

67. R v. Adams (*Adams II*), [1998] 1 Crim. App. 377 at 379 (Eng.).

68. *See Adams I*, 2 Crim. App. at 470 (discussing the defense estimate of the incidence of the group of nine genetic markers in the population as 1 in 2 million); *Adams II*, 1 Crim. App. at 379 (discussing defense claims that the sample size was too small and that the prosecution expert improperly used a pen to draw in a DNA band).

69. *See supra* notes 67–68.

70. *Adams I*, 2 Crim. App. at 477.

exclusionary non-matches: the victim's contrary description of the culprit and inability to identify Adams's and his 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.[71] 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.[72] The jury convicted again. On a second appeal, again alleging that the jury was improperly instructed, the Criminal Division of the Court of Appeal affirmed. This time the court banned the use of Bayes' Theorem in future criminal trials. In doing so, it validated dicta in its earlier decision that jurors 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."[73] 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."[74]

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* evidence of guilt was the statistically aggregated effect of a series of *matches* of non-unique genetic traits. Absent aggregative analysis—the very "mathematical formulas" to inform jury deliberation that the court rejected in the process of upholding Adams's conviction—the defendant could not have been convicted or even charged with a crime. 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 *Collins*, announced in 1968.[75] 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 indicating that there was a vanishingly small

71. *Id.* at 482.

72. *See* Donnelly, *supra* note 63, at 16.

73. *Adams I*, 2 Crim. App. at 481.

74. *Adams II*, 1 Crim. App. at 384 (quoting R v. Doheny, [1997] 1 Crim. App. 369 at 374 (Eng.) (quoting *Adams I*, 2 Crim. App. at 481)).

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

probability of finding a matching couple at random in Los Angeles and thus, it was suggested, a vanishingly small chance that there was any other such couple that could have committed the crime.[76]

The California Supreme Court overturned the resulting conviction.[77] In doing so, the court could have taken a narrower route, criticizing the prosecutors for inventing an excuse to emphasize the one, weak, and potentially prejudicial fact that 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;[78] 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);[79] and the expert asked the wrong question (how likely it is that a couple chosen 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).[80] Instead, the court went on to give three reasons why even properly implemented quantification was a bad idea: the jury would give undue weight to statistical evidence presented by an expert in numeric form;[81] such analysis was beyond the ken of the defense to understand and effectively rebut;[82] 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?"[83]

76. *Id.* at 36–37. 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. *Id.* at 34. Neither eyewitness could identify the defendants, Malcolm Collins and his wife, Janet, as the perpetrators. *Id.* at 36. Operating without data, the expert witness estimated frequencies for each trait of the perpetrator couple that matched the defendant couple (e.g., .001 for "[i]nterracial 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. *Id.* at 36–37 & n.10. The jury found the defendants guilty. *Id.* at 33.

77. *Id.*

78. *Id.* at 38.

79. *Id.* at 38–39.

80. *Id.* at 40; *see supra* note 76 and accompanying text.

81. *Collins*, 438 P.2d at 40–41 ("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 . . . .").

82. *Id.* at 41.

83. *Id.* at 40 (emphasis omitted) ("[N]o mathematical equation can prove beyond a reasonable doubt (1) that the guilty couple *in fact* possessed the characteristics described by the

In the decades since *Collins* was decided, it has been approvingly cited by courts in many American jurisdictions for the proposition that aggregating probabilities associated with different items of matching evidence is impermissible.[84] 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 fourth argument even more sweeping than the others.[85] 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.[86] Tribe's 1971 article extended this critique, which *Collins* had applied to the multiplication rule used there, to a proposal to avoid the foibles of the multiplication rule by using more sophisticated Bayesian analysis.[87] Since then, many other scholars have debated, and mainly endorsed and extended, Tribe's objections to the use of systematic aggregative analysis to establish identity in criminal process.[88]

---

People's witnesses, or even (2) that only *one* couple possessing those distinctive characteristics could be found in the entire Los Angeles area.").

84. *See, e.g.*, cases cited *infra* note 444.

85. Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329 (1971).

86. *See id.* at 1358–78; *cf. id.* at 1330 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 . . . .").

87. *See id.* at 1331–78 (criticizing the proposal in Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 HARV. L. REV. 489 (1970), to use Bayesian analysis to help resolve identity issues in criminal trials).

88. For a sampling of the voluminous literature, see generally PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE: THE USES AND LIMITS OF BAYESIANISM (Peter Tillers & Eric D. Green eds., 1988); Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 VA. L. REV. 1491 (2001); Ronald J. Allen, *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*, 65 TUL. L. REV. 1093 (1991); Lea Brilmayer & Lewis Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116 (1978); L. Jonathan Cohen, Dialogue, *Subjective Probability and the Paradox of the Gatecrasher*, 1981 ARIZ. ST. L.J. 627; Eric D. Green, Foreword, Symposium, *Probability and Inference in the Law of Evidence*, 66 B.U. L. REV. 377 (1986); David Kaye, *The Paradox of the Gatecrasher and Other Stories*, 1979 ARIZ. ST. L.J. 101; Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357 (1985); Daniel Shaviro, Commentary, *Statistical-Probability Evidence and the Appearance of Justice*, 103 HARV. L. REV. 530, 547–50 (1989); Peter Tillers, *Decision and Inference*, 13 CARDOZO L. REV. 253 (1991); articles cited *infra* note 444.

### B. OBJECTIONS TO AGGREGATIVE ANALYSIS OF MATCHES AND NON-MATCHES

Tribe's fourth critique can be subdivided into three parts: First, the use of aggregative methodology offends fundamental principles, such as the presumption of innocence[89] or the principle that we punish people based only on proof of what they did, not proof that they are members of a class.[90] 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.[91] 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.[92] 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.'"[93]

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,[94] 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

89. *See, e.g.*, Tribe, *supra* note 85, at 1355–67 (arguing that Bayesian analysis compromises the presumption of innocence by requiring jurors to consider the opening odds of guilt); *infra* notes 128–29 and accompanying text.

90. *See* Tribe, *supra* note 85, at 1359–61 (arguing that Bayesian analysis would compromise the right to be convicted based on individual, not class, evidence by permitting juries to convict a defendant even after acknowledging, say, a 3% chance that she is innocent); *see also* RONALD L. ALLEN ET AL., EVIDENCE: TEXT, PROBLEMS, AND CASES 164–65 (4th ed. 2006); sources cited *infra* notes 127–28 and accompanying text.

91. *See* Tribe, *supra* note 85, at 1335–37.

92. *See id.* at 1360–61.

93. People v. Louie, 205 Cal. Rptr. 247, 262 (App. Dep't Super. Ct. 1984) (quoting People v. Cella, 188 Cal. Rptr. 675, 684 (Ct. App. 1984)); *see also, e.g.*, L. Jonathan Cohen, *Can Human Irrationality Be Experimentally Demonstrated?*, 4 BEHAV. & BRAIN SCI. 317, 329 (1981); Tribe, *supra* note 85, at 1361–63.

94. For early criticism of the use of DNA in court, see *People v. Castro*, 545 N.Y.S.2d 985, 996–99 (Sup. Ct. 1989); Eric S. Lander, Commentary, *DNA Fingerprinting on Trial*, 339 NATURE 501 (1989); Stephen M. Patton, *DNA Fingerprinting: The* Castro *Case*, 3 HARV. J.L. & TECH. 223 (1990); Janet C. Hoeffel, Note, *The Dark Side of DNA Profiling: Unreliable Scientific Evidence Meets the Criminal Defendant*, 42 STAN. L. REV. 465 (1990); Peter J. Neufeld & Neville Colman, *When Science Takes the Witness Stand*, SCI. AM., May 1990, at 46.

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.[95] Older examples are eyewitness identifications matching the witness's memory of the culprit to a suspect and confessions or 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 many tiny lines and intersections found in a latent partial fingerprint at a crime scene match those taken from a suspect.[96] Lines and

95. *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 HASTINGS L.J. 1069, 1101 (1998).

96. *See, e.g.*, COMM. ON IDENTIFYING THE NEEDS OF THE FORENSIC SCI. CMTY. ET AL., NAT'L RESEARCH COUNCIL, STRENGTHENING FORENSIC SCIENCE IN THE UNITED STATES: A PATH FORWARD 53, 100, 141–44 (2009), *available at* https://www.ncjrs.gov/pdffiles1/nij/grants/228091.pdf (criticizing as not scientifically validated the longstanding use by fingerprint analysts of subjective evaluation of overall friction-ridge patterns, and calling for the aggregation of numerical scores for the many individual elements of latent fingerprints); Jennifer L. Mnookin et al., *The Need for a Research Culture in the Forensic Sciences*, 58 UCLA L. REV. 725, 731, 734–35 (2011) (reaching similar conclusions about fingerprint analysis and other pattern-identification disciplines); Laura Spinney, *The Fine Print*, 464 NATURE 344, 346 (2010) (reporting that some forensic experts believe that the only way to avoid false-positive fingerprint matches is for "fingerprint evidence [to] be interpreted in probabilistic terms," using data on "how fingerprint patterns vary across populations and how often various components or

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.[97] 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*,[98] 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.[99]

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."[100] As in Paley's case, this claim typically is made to refute the idea that "circumstantial" evidence is inferior to "direct" evidence.[101] 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.[102] 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

combinations of components crop up," for example, "a particular configuration of bifurcations, ridge endings and the like [is] found in 40% of a given population"); *infra* note 285 and accompanying text.

97. *See* Paul C. Giannelli, *Forensic Science: Why No Research?*, 38 FORDHAM URB. L.J. 503, 513–18 (2010) (arguing that the *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).

98. *See, e.g.*, United States v. Mack, 258 F.3d 548, 554 (6th Cir. 2001) ("[S]tandard 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 . . . .").

99. *See* RICHARD O. LEMPERT ET AL., A MODERN APPROACH TO EVIDENCE: TEXT, PROBLEMS, TRANSCRIPTS AND CASES 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 †.

100. WILLIAM PALEY, THE PRINCIPLES OF MORAL AND POLITICAL PHILOSOPHY 932 (Liberty Fund 2002) (1785).

101. *See infra* Part V.C.1 (discussing judicial definitions of "direct" and "circumstantial" evidence that suggest a preference for the former).

102. *See* Simon A. Cole, *Forensics Without Uniqueness, Conclusions Without Individualization: The New Epistemology of Forensic Identification*, 8 LAW PROBABILITY & RISK 233, 234–37 (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).

conceptually are on the "circumstantial" side of the divide, instead qualify as presumptively stronger "direct" or "unique" evidence.[103]

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.[104] 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 happened.[105] In these examples, as in the more obviously "circumstantial" ones described

103. *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. INNOVATION 231, 253–58 (2008) (describing studies showing that mock jurors instructed on the difference between direct and circumstantial evidence often mischaracterize highly probative circumstantial evidence as "direct"); *infra* notes 364–67 and accompanying text.

104. *See, e.g.*, Sylvester v. SOS Children's Vills. Ill., Inc., 453 F.3d 900, 903 (7th Cir. 2006) ("[A]ll evidence, even eyewitness testimony, requires drawing inferences; the eyewitness is drawing an inference from his raw perceptions. '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.'" (quoting Milam v. State Farm Mut. Auto. Ins. Co., 972 F.2d 166, 170 (7th Cir. 1992))); DePass v. United States, 721 F.2d 203, 207 (7th Cir. 1983) (Posner, J., dissenting) (arguing that "almost all legal evidence is probabilistic" and that "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").

105. There is probative value in a confession without details ("I did it") or a lineup 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.*, Steven A. Drizin & Richard A. Leo, *The Problem of False Confessions in the Post-DNA World*, 82 N.C. L. REV. 891, 919–20 (2004) (noting that innocent suspects sometimes confess to stop exhausting interrogations, to please authorities, or because of youth or mental disability); Stephen Greenspan, *There Is More to Intelligence than IQ*, *in* CONVICTING THE INNOCENT: THE STORY OF A MURDER, A FALSE CONFESSION, AND THE STRUGGLE TO FREE A "WRONG MAN" 148 (Donald S. Connery ed., 1996) (documenting how a suspect's youth, inexperience with the justice system, mental illness, and suggestibility disposed him to confess falsely); Sarah Anne Mourer, *Reforming Eyewitness Identification Procedures Under the Fourth Amendment*, 3 DUKE J. CONST. L. & PUB. POL'Y 49, 56–57 (2008) (discussing disposition of eyewitnesses to make inaccurate identifications to provide closure).

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.

Although our intuitions tell us that some evidence is powerful because it reveals a match between a unique trait of the culprit and a suspect, the power of the evidence is almost never due in fact to the uniqueness of the evidential trait. The evidence 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 of non-unique traits increases, indeed multiplies, the probability of guilt—and similarly, we will argue, the power of each additional non-match multiplies the probability of innocence. We will 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.[106] A central goal of these proposed improvements is to force into the open as many individual details as possible, to permit an 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 verify that investigators did not feed details to the suspect.[107] Similarly, proposals to document pre-lineup eyewitness descriptions with sketch artists, and to cross-reference them to descriptions other witnesses independently gave, help to specify and clarify the many separate features that eyewitness identifications typically conflate, exposing the number and weight of each match and non-match.[108] 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.[109]

106. *See, e.g.*, GARRETT, *supra* note 36, at 14–83 (cataloguing confessions and identifications later shown to be false); Gisli H. Gudjonsson, *False Confessions and Correcting Injustices*, 46 NEW ENG. L. REV. 689, 690–91 (2012).

107. For articles advocating the videotaping of interrogations, see *supra* note 41; *infra* notes 509–10.

108. *See* Neil Vidmar, James E. Coleman, Jr. & Theresa A. Newman, *Rethinking Reliance on Eyewitness Confidence*, 94 JUDICATURE 16, 17 (2010). For parallel proposals in the fingerprint context, see *supra* note 96 and accompanying text.

109. *See supra* note 41 (noting the increasing number of jurisdictions that videotape confessions).

### 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*,[110] 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,"[111] appropriating the common misperception of fingerprint evidence as unique to a single person,[112] inculpatory DNA evidence does not depend on any unique feature of individuals. 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.[113]

Modern DNA profiling uses a Polymerase Chain Reaction ("PCR"), followed by gel or capillary electrophoresis, to identify DNA sequences known as short tandem repeats ("STRs").[114] 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 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 number of CATGs in a row) varies among individuals. One person may have five repeats, another may have six, and so on. Each different number of repeats is called an "allele."[115]

110. *See supra* notes 75–83 and accompanying text.

111. *See, e.g.*, Lander, *supra* note 94, at 505; William C. Thompson & Simon Ford, *DNA Typing: Acceptance and Weight of the New Genetic Identification Tests*, 75 VA. L. REV. 45, 49 (1989).

112. *See supra* note 96.

113. Although, with the exception of monozygotic twins, each individual has a unique genome, forensic DNA profiling typically cannot examine enough of an individual's entire genome to establish uniqueness. *See* HARLAN LEVY, AND THE BLOOD CRIED OUT: A PROSECUTOR'S SPELLBINDING ACCOUNT OF THE POWER OF DNA 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.").

As we note above, the idea that evidence needs to be aggregated to reveal its probative value is not new to DNA and has been recognized for years, for example, in fingerprint analysis and paternity cases. *See supra* notes 96–101 and accompanying text. We use DNA to make the point because of its widely accepted status as the gold standard of evidence.

114. DAVID H. KAYE, THE DOUBLE HELIX AND THE LAW OF EVIDENCE 178–91 (2010).

115. MICHAEL J. SAKS ET AL., ANNOTATED REFERENCE MANUAL ON SCIENTIFIC EVIDENCE SECOND 676 (2005–2006 ed. 2000) (noting that each individual has two alleles at a given STR locus, one from each chromosome).

Variation in the alleles found at each STR locus can help discriminate among individuals.[116]

Individually, however, each allele is fairly common. In some cases, 25% of the relevant population share a given allele.[117] 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.[118] 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.[119] 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 FBI[120]), and if the incidence of each STR is 20% (making each trait more common than left-handedness and male-pattern 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,[121] as the jury will be informed.

Going beyond the multiplication rule, several U.S. courts in civil and criminal paternity cases have endorsed Bayes' Theorem as the most acceptable way to update the "prior odds" estimate of paternity 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

116. KAYE, *supra* note 114, at 188.

117. *See, e.g.*, John M. Butler et al., *Allele Frequencies for 15 Autosomal STR Loci on U.S. Caucasian, African American, and Hispanic Populations*, 48 J. FORENSIC SCI. 908, 910–11 (2003).

118. *See* Neufeld & Colman, *supra* note 94, at 50 ("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 at every site] becomes increasingly unlikely . . . ."); *see also* Desmond C.C. Gan & Rodney D. Sinclair, *Prevalence of Male and Female Pattern Hair Loss in Maryborough*, 10 J. INVESTIG. DERMATOL. SYMP. PROC. 184, 185–86 (2005) (reporting that approximately 17% of men in their twenties exhibited symptoms of male-pattern baldness); Michel Raymond et al., *Frequency-Dependent Maintenance of Left Handedness in Humans*, 263 PROC.: BIOLOGICAL SCI. 1627, 1627 (1996) (estimating frequency of left handedness as 10–13%).

119. In DNA analysis, the researcher estimates "the frequency of the alleles at each locus (the 'single-locus genotypes'), and then [via the 'Basic Product Rule'] combines those figures to estimate the frequency of the combination of single-locus genotypes (the 'multiple single-locus' or 'multilocus genotype')." KAYE, *supra* note 114, at 89.

120. *Id.* at 189.

121. *Id.* at 187–88.

profile if someone else is his father).[122] 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.[123]

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 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, or ear-lobe configuration. By analogy to the DNA paternity cases, both Bayes' Theorem and 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 biostatistician will represent the aggregate strength of the string of matches as a number—the product of the probabilities associated with each matching allele. In other words, an expert 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 allowed. Even if the data 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 from consulting outside of court, a statistician to report a number.

Indeed, with the exception of testimony about the probability of a random match of forensic traits such as blood types and DNA profiles, which the courts have grown increasingly likely to allow,[124] the general rule is to

122. *See, e.g.*, T.A.T. v. R.E.B. (*In re* Paternity of M.J.B.), 425 N.W.2d 404, 408–09 (Wis. 1988) (describing various statistical techniques).

123. *See* 4 FAIGMAN ET AL., *supra* note 47, §§ 31:49–50 (describing use of Bayes' Theorem in paternity cases); DAVID H. KAYE ET AL., THE NEW WIGMORE: EXPERT EVIDENCE § 12.4.3 (2004) (same); *infra* text accompanying notes 175–79 (same).

124. *See, e.g.*, People v. Mountain, 486 N.E.2d 802, 804–06 (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?*,

exclude testimony about the frequency of identifying traits. The result is to preclude 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, most courts forbid jurors to consider the study. For example, courts may bar evidence in insurance-fraud cases of how often particular medical diagnoses are made; 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.[125] As Professor Jonathan 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.[126]

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' claims that quantification and aggregation threaten the foundations of our criminal justice system.[127] 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.[128] The verdict must

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, 1612 (2001) (observing that "[m]ost criminal courts . . . allow the presentation of the random match statistic" (i.e., the frequency of a forensic trait in the general population)).

125. *See* Koehler, *supra* note 124, at 391–92 n.106 (citing cases). *But see id.* at 389–90 (citing decisions allowing statistical evidence of (1) the small probability of Sudden Infant Death Syndrome, (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 (3) the small probability of a particular pattern of accurate and inaccurate answers to prove cheating on a test).

126. *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 187. DNA evidence thus arguably cannot satisfy the "reference class" requirements that bar aggregative analysis of other evidence.

127. *See supra* notes 86–93 and accompanying text.

128. *See* Nesson, *supra* note 88, at 1378 ("[A] probable verdict may not be acceptable [to the public], and an acceptable verdict may not be probable."); Tribe, *supra* note 85, at 1372–75 (claiming that probabilistic evidence "dehumanizes" justice and weakens public and party support for the legal system).

convey to the public the message that the defendant *is* guilty, not that she "probably" or "almost certainly" is.[129]

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 the jurors may find the defendant guilty though they are "fully aware" that their verdict is based on "probabilities" and "may be mistaken."[130] 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 the statistical realities of DNA evidence.[131] One such reality is that DNA evidence 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 and cannot reveal the probability of guilt.[132] Second, there are a number of reasons why the probability that the defendant is guilty is always *lower* than the probability that the defendant was the source of the biological material, including the existence of a probability greater than zero that investigators erred in collecting or analyzing the evidence or that there was an innocent reason why the defendant's or victim's genetic material was found in the incriminating location.[133] Despite these realities, courts on both sides of the Atlantic have no problem upholding convictions based on little more than a DNA match that actually is a series of mathematically aggregated matches.[134]

129. *See* Nesson, *supra* note 88, 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 85, at 1372–73 (worrying that explicitly premising convictions on evidence of probabilities less than one in criminal trials could undermine the policy of acquittal in the face of reasonable doubt).

130. *See, e.g.*, Victor v. Nebraska, 511 U.S. 1, 18, 22–23 (1994) (upholding constitutionality of a jury instruction that included this language).

131. McDaniel v. Brown, 130 S. Ct. 665 (2010).

132. *See id.* at 670; *infra* notes 169–70 and accompanying text.

133. *See* Norman Fenton & Martin Neil, *Avoiding Probabilistic Reasoning Fallacies in Legal Practice Using Bayesian Networks*, 36 AUSTL. J. LEGAL PHIL. 114, 122–23 (2011) ("Errors in the DNA typing process can result in a reported match where there is no 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 scene." (footnote omitted)); *infra* notes 171–74 and accompanying text. *See generally* Jonathan J. Koehler, *Error and Exaggeration in the Presentation of DNA Evidence at Trial*, 34 JURIMETRICS J. 21 (1993).

134. *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 of semen at the

Clearly, our criminal justice system exhibits no blanket preference for a false certainty over estimated probabilities as a matter of liberal-democratic fundaments, and it should forbear applying any such preference 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 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.[135] That cognitive bias, however, can lead us to accept demonstrably false things as true. The bias 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.[136] 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.[137] 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*

The question remains, however, whether systematically aggregated probabilities associated with non-exclusionary non-matches can actually help reach accurate answers, and not simply confuse or mislead 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.[138] In view of how often exonerations highlight potentially identifying traces found during police investigations that did *not* match the wrongly convicted defendant but did match the actual

crime scene and on testimony that the probability of a random match was one in 135 million). Spencer was executed on April 27, 1994. *See Searchable Execution Database*, DEATH PENALTY INFO. CENTER, http://www.deathpenaltyinfo.org/executions (last visited Oct. 2, 2012); *see also supra* notes 63–74 and accompanying text (discussing *R v. Adams*).

135. *See infra* Part V.A.3.c.

136. *See infra* notes 256–58 and accompanying text.

137. *See generally* GERD GIGERENZER, RECKONING WITH RISK: LEARNING TO LIVE WITH UNCERTAINTY (2002).

138. *See supra* notes 110–21 and accompanying text.

perpetrator,[139] 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 probabilities large enough 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.[140] 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, the use of DNA evidence to establish criminal identity and paternity initially raised these same concerns, but over time the adversarial system worked hand-in-glove with the relevant technologists (biologists and statisticians) to allay the concerns. 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.[141]

Although celebrated today, inculpatory DNA analysis initially was crude and controversial. In *People v. Castro*,[142] 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,[143] validated other prosecution testimony,[144] including that of his colleague Dr. Michael Baird, which placed "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."[145] Testifying for the defense, Dr. Eric Lander, a Harvard and MIT mathematician and scientist and a MacArthur award recipient,[146] "examined the same data and arrived at odds of one in 24."[147] After appearing as opposing experts, Drs. Roberts and Lander met privately while the case was in recess, decided that 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 data in

139. *See supra* notes 42–44 and accompanying text.

140. *See supra* notes 75–83 and accompanying text.

141. *See* KAYE, *supra* note 114, 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).

142. People v. Castro, 545 N.Y.S.2d 985 (Sup. Ct. 1989).

143. *See* LEVY, *supra* note 113, at 42 (describing the *Castro* trial).

144. *Id.* at 43.

145. *Castro*, 545 N.Y.S.2d at 985.

146. *See* LEVY, *supra* note 113, at 43; *see also* Gina Kolata, *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 (profiling Dr. Eric Lander).

147. Neufeld & Colman, *supra* note 94, at 48.

this case are not scientifically reliable enough to support the assertion that the samples . . . do or do not match."[148] In a carefully reasoned decision in August 1989, the trial judge concluded that DNA matching is a potentially reliable form of identity evidence but that some of the DNA evidence presented was inadmissible in the case due to a number of methodological improprieties in data mining and statistical analysis.[149] 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 similar problems in the future.[150] A 1990 *Scientific American* article by one of Castro's lawyers, Peter Neufeld, presented an overlapping set of critiques and methodological and procedural solutions.[151]

The defects identified in early cases and commentaries fell into four categories. First, 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.[152] Second, 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.[153] Third, technicians used invalid statistical methods to determine the chance of a random match, including (1) the failure to sample enough human subjects to generate reliable frequencies of

148. Lander, *supra* note 94, at 504 (internal quotation marks omitted); *see* LEVY, *supra* note 113, at 47.

149. *Castro*, 545 N.Y.S.2d at 996–97, 999 (criticizing the prosecution's DNA evidence for failing (1) to use proper probes when analyzing degraded biological samples; (2) to include male and female controls when testing for sex of sample on the defendant's watch; and (3) to test for non-human DNA in samples, rather than assuming all strands were of human origin).

150. *See* Lander, *supra* note 94, 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; and reliance on assumptions about population genetics that, for example, ignore heterogeneity in Hispanic populations).

151. *See, e.g.*, Neufeld & Colman, *supra* note 94, at 50–53 (advocating (1) improved methods to determine whether an allele from one sample is the same as that from another, especially with small samples that prevent retesting; (2) avoiding misjudgments due to contamination of crime-scene specimens, bacterial degradation, improperly prepared gels, and over-concentrated samples without available control groups; (3) the standardization of methods for sampling frequencies and population estimates across laboratories; and (4) studies to determine whether differences in allele frequency among ethnic subgroups invalidate statisticians' assumption of random mating across racial groups).

152. *See, e.g.*, *Castro*, 545 N.Y.S.2d at 993–94; KAYE, *supra* note 114, at 51; Lander, *supra* note 94, at 502–03; Neufeld & Colman, *supra* note 94, at 51.

153. *See, e.g.*, *Castro*, 545 N.Y.S.2d at 994; Lander, *supra* note 94, at 502–03; Neufeld & Colman, *supra* note 94, at 51; Hoeffel, *supra* note 94, at 479.

particular alleles in different populations;[154] (2) 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, which creates 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);[155] (3) 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);[156] and (4) the failure of private laboratories to reveal "proprietary" assumptions about the frequency of alleles in different populations, preventing cross-laboratory comparison and peer review.[157] Fourth, laboratories failed to document instances and patterns of shoddy or improper techniques in collecting, handling, and analyzing samples.[158]

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 Sciences, and other independent bodies.[159] 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,[160] 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 was based on an unproven assumption that particular chemical probes would isolate material at only a single locus on the crime scene and the suspect's samples and, thus, that the radiographs taken of the two samples could be compared to

154. *See* Patton, *supra* note 94, at 236 n.41.

155. *See, e.g.*, *Castro*, 545 N.Y.S.2d at 992; KAYE, *supra* note 114, at 124; Lander, *supra* note 94, at 504; Patton, *supra* note 94, at 236; Hoeffel, *supra* note 94, at 489–90.

156. *See, e.g.*, People v. Pizarro, 10 Cal. App. 4th 57, 93–94 (Ct. App. 1992); D.H. Kaye, *Logical Relevance: Problems with the Reference Population and DNA Mixtures in* People v. Pizarro, 3 LAW PROBABILITY & RISK 211, 214 (2004).

157. *See, e.g.*, People v. Wesley, 633 N.E.2d 451, 463–64 (N.Y. 1994) (Kaye, C.J., concurring).

158. *See, e.g.*, *Castro*, 545 N.Y.S.2d at 997; KAYE, *supra* note 114, at 101; Hoeffel, *supra* note 94, at 493–95; Neufeld & Colman, *supra* note 94, at 53.

159. *See* COMM. ON DNA TECH. IN FORENSIC SCI. ET AL., NAT'L RESEARCH COUNCIL, DNA TECHNOLOGY IN FORENSIC SCIENCE (1992) [hereinafter NRC I]; COMM. ON DNA TECH. IN FORENSIC SCI.: AN UPDATE, NAT'L RESEARCH COUNCIL, THE EVALUATION OF FORENSIC DNA EVIDENCE (1996) [hereinafter NRC II].

160. *See* LEVY, *supra* note 113, 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* note 94.

see if they matched. Responding to defense attorneys' and judicial concerns, analysts more clearly established that the chemical probes used 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.[161] Additionally, more automated and repetitive techniques require less genetic material and eliminate much of the subjectivity associated with previous methods for determining whether the gene sequences in the two pictures were sufficiently distinct and similar in length to establish a match, permitting independent replication in more cases.[162]

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.[163] A variety of sophisticated methods of adjusting random-match probabilities based on the size of the database of reference samples have also been developed.[164]

Interdependence worries based on the possibility that particular alleles and clusters of alleles 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. As a result, the actual aggregate probability of a random-match was likely to be much lower in fact than the conservative probability offered in court.[165] 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."[166]

161. *See, e.g.*, KAYE, *supra* note 114, at 50.

162. *See id.* at 178–80.

163. *See id.* at 138.

164. *See, e.g.*, JOHN M. BUTLER, FUNDAMENTALS OF FORENSIC DNA TYPING 244–49 (2010).

165. *See* NRC I, *supra* note 159, at 82–85 (concluding that ceiling frequencies are a scientifically acceptable alternative to sampling every conceivable subpopulation); NRC II, *supra* note 159, at 5–8 (identifying database alternatives when a particular subpopulation has no available sample set).

166. Koehler, *supra* note 124, at 394. Professor Koehler noted 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." *Id.* at 393; *see also* Eric S.

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, 1%) into the probability of the *defendant's guilt* (99%).[167] This transposition ignores the prior odds. Even if 99% 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%, 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 briefcase carrier is a lawyer, we must also know the

> a priori 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 a priori likelihood that the defendant in question is guilty.[168]

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 prior estimate of the probability of a suspect's guilt after receiving [DNA] evidence accompanied by incidence rate statistics."[169] 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 potentially the routine use of Bayes' Theorem to highlight the role of prior odds.[170]

Bayes' Theorem helps formalize two related problems. Properly calculating the likelihood-ratio denominator—the probability that a match

---

Lander & Bruce Budowle, *DNA Fingerprinting Dispute Laid To Rest*, 371 NATURE 735 (1994) (similar).

167. *See* McDaniel v. Brown, 130 S. Ct. 655, 670 (2010) ("[I]f a juror is told the probability a member of the general population would share the same DNA is 1 in 10,000 (random match probability), 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 (source probability), then he has succumbed to the prosecutor's fallacy."); William C. Thompson & Edward L. Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 LAW & HUM. BEHAV. 167, 169–71 (1987).

168. Thompson & Schumann, *supra* note 167, at 170.

169. *Id.* at 170 n.2.

170. *See McDaniel*, 130 S. Ct. at 665, 670–76 (acknowledging that a prosecutor's resort to the prosecutor's fallacy in closing argument could mislead jurors, potentially making the trial "fundamentally unfair," but denying relief because the issue had not been properly raised below); Christopher M. Triggs & John S. Buckleton, Comment, *Why the Effect of Prior Odds Should Accompany the Likelihood Ratio When Reporting DNA Evidence*, 3 LAW PROBABILITY & RISK 73, 76 (2004).

would appear though the defendant is innocent—requires consideration not only of the random-match probability but also (1) the frequency of laboratory false positives[171] 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.[172] Courts now address the first part of the problem by admitting evidence of error rates for particular laboratories and analysts.[173] 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 reasons for the presence of the defendant's DNA may motivate solutions to the rest of the problem, again potentially including use of Bayesian analysis to highlight the key considerations.[174]

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 revolutionized proof of paternity,[175] including proof of identity in rape cases involving minors or severely disabled nursing-home patients who give birth to a child.[176] 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 (1) 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;[177] (2) experts for each side to propose prior odds based on

171. *See* Koehler, *supra* note 124, at 394 & n.118 (citing laboratory error rates ranging from 1 out of 67 to 1 out of 345 false positives).

172. *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, 431 (1997).

173. *See* Koehler, *supra* note 124, at 394–95 & nn.119, 124 (citing cases).

174. *McDaniel*, 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 also* articles cited *supra* notes 167–72 (advocating full use of Bayesian analysis to assure the accuracy of inferences juries draw from DNA evidence).

175. *See supra* notes 122–23 and accompanying text.

176. *See, e.g.*, Griffith v. State, 976 S.W.2d 241, 245–46 (Tex. Ct. App. 1998).

177. *See, e.g.*, Davis v. State, 476 N.E.2d 127, 138 (Ind. Ct. App. 1985) (affirming conviction premised on proof of the defendant's parental relationship to the victim, which was established using Bayes' Theorem and a 50–50 estimate of the prior odds); Kammer v. Young, 535 A.2d 936, 940–42 (Md. Ct. Spec. App. 1988) (permitting an expert's estimate of 50–50 prior odds, and inviting the defendant to offer evidence and argument attacking that assumption); Commonwealth v. Beausoleil, 490 N.E.2d 788, 797 n.19 (Mass. 1986) (similar); *Griffith*, 976 S.W.2d at 245–46 (approving use of an estimate of 50–50 prior odds in a case involving a

their understanding of the evidence and letting the jury choose between the competing estimates or make its own;[178] and (3) 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.[179]

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.[180] Statistical hurdles that have been overcome include several that *Collins* and Tribe treated as nearly insurmountable obstacles to systematically aggregating probabilities associated with identifying traits.[181] 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.[182] 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.

---

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 90, at 163 (criticizing the 50–50 assumption and "assessment of the odds of guilt or liability *before* the receipt of . . . evidence").

178. *See, e.g.*, T.A.T. v. R.E.B. (*In re* Paternity of M.J.B.), 425 N.W.2d 404, 409–11 (Wis. 1988) (rejecting the 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).

179. *See, e.g.*, State v. Spann, 617 A.2d 247, 253–54 (N.J. 1993) (ruling that an expert witness testifying to probabilities in a criminal paternity case must list and explain the effect of different prior probabilities); M. v. Marvin S., 656 N.Y.S.2d 802, 809 (Fam. Ct. 1997) (affirming a paternity judgment that was based on testimony by an expert who "utilized the 'chart approach' . . . in showing the effect of a range of Prior Probability values on the actual genetic test results"); Plemel v. Walter, 735 P.2d 1209, 1219 (Or. 1987) (requiring a paternity expert to identify outcomes based on prior odds estimates ranging from 0 to 100%); *see also* KAYE ET AL., *supra* note 123, § 12.4.3, at 493–95 (endorsing the 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. LEGAL STUD. 395, 403, 429–36 (2005) (presenting the results of an empirical study showing that the "chart format" reduces error by mock jurors using Bayesian analysis).

180. *See* KAYE, *supra* note 114, at 190–91 (describing the courts' increasing impatience with challenges to forensic DNA probabilities).

181. *See supra* notes 84–88 and accompanying text; *see also* Tribe, *supra* note 85, at 1365 (predicting that mathematical analysis, even if based on accurate calculations of probabilities and statistical odds, would make little headway in courts).

182. *See supra* notes 163–66, 175–79 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 to determine objectively the length of the tandem repeats at each locus, providing strong bases for objection in court. It was only through the adversary system that those objections motivated the invention of more exacting—if still imperfect—probes and other analyses.[183] Similar objections are beginning to force parties offering fingerprint evidence to acknowledge its status and improve its performance as a method for systematically aggregating the effects of multiple individually inconclusive matches.[184] The next Subpart 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 who committed 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 that population), and balding (27% of that population). The perpetrator is among the subset of the population that secretes blood-group antigens into bodily fluids such as semen (80% of the population), and the antigens secreted into semen found in the rape victim reveal that the perpetrator's blood type is A-negative (6% of the relevant population). Police receive an anonymous call from someone claiming that an acquaintance—who is only described as a white male currently standing alone on the northeast corner of Fifth and Main Streets—told her that he just raped a woman. The caller says she knows nothing about where the rape occurred or who the victim might be. 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 a white adult male encountered "at random" (i.e., for reasons independent of the traits being matched) would be green-eyed, 5′3″ or shorter, left-handed, bald, and an A-negative secretor is estimated by multiplying the frequency of the identifying traits: *.09 x .13 x .05 x .27 x .06 x .8*, resulting in a probability of just under 8 in 1,000,000.

183. *See supra* notes 161–62 and accompanying text.

184. *See supra* note 96 and accompanying text.

Because information on these traits has been collected for years, their use in this manner avoids 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.[185] 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.[186] Forty years after *Collins*, however, criticism of data mining runs in the opposite direction: we collect too much.[187] 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—in the population at large, and for the factors in the rape and *Collins* cases, 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-pattern baldness, cigarette smoking, car makes and models, eye color, size, weight, and many others.[188] 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 and behaviors, including patterns of dress, tastes, habits, and other preferences, much of which can be segmented by state, city, and postal code.[189]

---

185. *See* Tribe, *supra* note 85, at 1335 (criticizing the prosecution's probabilistic testimony in *Collins* as devoid of empirical support for the probabilities used).

186. *See* KAYE, *supra* note 114, at 88 (citing early criticism of DNA evidence based on inadequate DNA data sets using small samples from FBI recruits).

187. For civil liberties criticisms of government collection of DNA samples, see, for example, Michael T. Risher, *Racial Disparities in Databanking of DNA Profiles*, GENEWATCH, July–Aug. 2009, at 22; Ben Protess, *The DNA Debacle: How the Federal Government Botched the DNA Backlog Crisis*, PROPUBLICA (May 5, 2009), http://www.propublica.org/article/the-dna-debacle-how-the-federal-government-botched-the-backlog-crisis-505; Jeffrey Rosen, *Genetic Surveillance for All*, SLATE (Mar. 17, 2009), http://www.slate.com/articles/news_and_politics/jurisprudence/2009/03/genetic_surveillance_for_all.html.

188. *See, e.g.*, *supra* note 118.

189. United States v. Jones, 132 S. Ct. 945, 963 (2012) (Alito, J., concurring in the judgment) (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 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 957 (Sotomayor, J., concurring) (noting the frequency with which people now disclose phone numbers, URLs, e-mail addresses, books, groceries, and medication purchases to cellular and Internet service providers and online retailers). For a small sampling of the burgeoning literature, see generally 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,

Law-enforcement agencies themselves diligently collect information capable of revealing the frequency of human traits.[190] 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 databases and are of sufficiently high resolution to distinguish faces; the color, make, and design of clothing; and automobiles.[191] 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.[192] Police officers in Oakland, California, use body-mounted personal-video cameras to record crimes, arrests, and traffic-stops.[193] 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 . . . computer[s] then compare[] . . . to a so-called 'hot list' of information on such things as stolen vehicles and other violations."[194] Law-enforcement

STRUCTURE, AND USAGE (2007); ROBERT NISBET ET AL., HANDBOOK OF STATISTICAL ANALYSIS AND DATA MINING APPLICATIONS (2009).

190. *See, e.g.*, COLLEEN MCCUE, DATA MINING AND PREDICTIVE ANALYSIS: INTELLIGENCE GATHERING AND CRIME ANALYSIS (2007); Monica C. Holmes et al., *Data Mining and Expert Systems in Law Enforcement Agencies*, 8 ISSUES INFO. SYS. 329 (2007); Hsinchun Chen et al., *Crime Data Mining: A General Framework and Some Examples*, COMPUTER, Apr. 2004, at 50.

191. *See, e.g.*, Al Baker, *Camera Scans of Car Plates Are Reshaping Police Inquiries*, N.Y. TIMES (Apr. 11, 2011), http://www.nytimes.com/2011/04/12/nyregion/12plates.html [hereinafter Baker, *Camera*] (discussing London's and New York City's systems of security cameras linked to police coordination centers).

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

193. Erica Goode, *Video, a New Tool for the Police, Poses New Legal Issues, Too*, N.Y. TIMES (Oct. 11, 2011), http://www.nytimes.com/2011/10/12/us/police-using-body-mounted-video-cameras.html?pagewanted=all.

194. Al Baker, *License Plate Reading Cameras Gain Legal Backing*, N.Y. TIMES (Apr. 14, 2011), http://cityroom.blogs.nytimes.com/2011/04/14/license-plate-reading-cameras-gain-legal-backing/. 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 [Zeledon'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.

officials have long mined traveler data to construct drug courier profiles, such as methods of paying for airline tickets, "source" and "destination" cities, duration of stay, forms of baggage, and modes of dress. Now, they also trawl airline records, checkpoints, and photographs for passengers who meet the predefined criteria, and the resulting evidence is often admitted in court.[195]

Of course, these activities raise serious civil liberties issues and are subject to proprietary, privacy, and security limitations on information sharing.[196] 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 a 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 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 p.m. in April.[197] 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.[198] In conjunction with other information markets,

---

Baker, *Camera*, *supra* note 191; *see also Jones*, 132 S. Ct. at 955 (Sotomayor, J., concurring) ("GPS monitoring generates a precise, comprehensive record of a person's public movements that reflects a wealth of detail about [their] . . . associations."); Heather Kelly, *Police Embracing Tech That Predicts Crimes*, CNN (July 9, 2012), http://www.cnn.com/2012/07/09/tech/innovation/police-tech/index.html (describing a computer algorithm in use by police departments in the U.S. that uses "locations of previous crimes, combined with sociological information about criminal behavior and patterns" to predict and work to prevent crime in 500-by-500 foot zones citywide, and also describing a process that involves using "microphones positioned around . . . cit[ies] to detect gunshots and triangulate their location within 40 to 50 feet" to alert police of the location).

195. *See, e.g.*, United States v. Sokolow, 490 U.S. 1, 10 (1989) (affirming reasonableness of search premised on the 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).

196. *See, e.g.*, *Jones*, 132 S. Ct. at 947, 954 (holding that, absent a warrant, officers' attachment of a GPS tracking device to Jones's car and tracing his movements on public streets 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 (2008).

197. *See, e.g.*, Family Educational Rights and Privacy Act (FERPA) of 1974, 20 U.S.C. § 1232g (2006) (barring release of data with individual identifiers but permitting release, for research and other purposes, of aggregate data from which individual identifiers have been removed); Health Insurance Portability and Accountability Act (HIPAA) of 1996, 42 U.S.C. § 1320d (2006) (similar).

198. For example, the *International Journal on Document Analysis and Recognition* "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." *International Journal on Document Analysis and Recognition*, SPRINGER, http://www.springer.com/computer/image+processing/journal/10032 (last visited Oct. 11, 2012).

demand by police, forensic laboratories, prosecutors, defense lawyers,[199] and courts surely could induce entrepreneurs and agencies to mine existing data sources and create new ones to generate frequencies of detectible human attributes in given reference populations. Indeed, Target Corporation, the national retailer, has created offices and laboratories in Minneapolis and Las Vegas that offer just these services.[200]

Some will surely object that these methods lack sufficient acuity or standards to distinguish videotaped, photographed, or otherwise recorded traits reliably to 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.[201] 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 a way to estimate them reliably.

---

199. Both prosecutors and defense lawyers are already taking advantage of related possibilities. *See, e.g.*, 2009 NAT'L ASS'N JUSTICE INFO. SYS. CONFERENCE AGENDA, *available at* http://najis.org/presentations/2009 (access by downloading zip file and opening "2009 agenda" file); *Forensics 411: How To Find Byte Marks*, INNOCENCE PROJECT OF MINN., http://www.ipmn.org/forensics_411_2010.php (last visited Oct. 11, 2012). Topics at the latter legal-training program included "recovering lost clusters of the digital DNA files . . . from a crime labs computers [sic]"; "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." *Id.*

200. *See* Press Release, Target Corp., Forensic Overview (Apr. 14, 2009), *available at* 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' "[a]ccreditation for digital evidence (video analysis, audio analysis and image analysis) and latent fingerprint analysis"); Target Corp., Presentation at National Association for Justice Information Systems Conference: Intelligence: Gathered, Analyzed, Disseminated & Managed (Sept. 15–17, 2009), *available at* http://najis.org/presentations/2009 (access by downloading zip file and opening "NAJIS Handout Sept. 15-17 2009[1] target" file); *see also* Brandt Williams, *Target's State-of-the-Art Forensics Lab Catches More than Just Shoplifters*, MINN. PUB. RADIO (Oct. 21, 2011), *available at* http://minnesota.publicradio.org/display/web/2011/10/21/target-forensics-lab/ (discussing the details of Target's Minneapolis-area forensics lab).

201. *See, e.g.*, United States v. Shabazz, 564 F.3d 280, 286–87 (3d Cir. 2009) (admitting testimony on whether the defendant did or did not appear in surveillance footage); Washington v. State, 961 A.2d 1110, 1114–19 (Md. 2008) (treating admission of compiled images from surveillance videotapes as non-harmless error, given the lack of authentication); *supra* notes 142–84 and accompanying text.

### *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 at least as likely to underuse aggregative analysis as the courts.[202] Unlike the parties in criminal trials, police detectives are not legally constrained in their use of statistical techniques in their investigations.[203] Yet, police often do not notice, much less capture, 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 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 turned out to be "small" *non*-matches.[204] Notwithstanding our own and the wider 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 an apparently 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 suggesting 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 might escape notice absent systematic aggregation.[205] In this Part, we explain why, in the absence of systematic aggregation, it is likely that powerful cognitive, structural, and legal forces will lead actors in the criminal justice process to undervalue "small" non-match evidence and risk convicting the innocent.

202. *See supra* Figure 2.

203. For example, good police practice calls for investigators to lay down a fine-meshed grid at crime scenes, then to videotape and meticulously examine the contents of each cell. *See infra* note 499.

204. *See infra* notes 329–57, 396–407 and accompanying text.

205. *See supra* notes 59–62 and accompanying text.

### *A. COGNITIVE RESISTANCE*

#### 1. Heuristic Economization

Human heuristic biases are unconscious cognitive tendencies to oversimplify the evaluation of uncertain probabilities in all facets of decision making. Such biases help explain the overvaluation of "big" matches and undervaluation of "small" non-matches.[206] Starting with the work of 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.[207]

#### 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.[208] Imagine a police detective assessing the probability 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

206. *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* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 463, 464–65 (Daniel Kahneman et al. eds., 1982) ("[H]euristics . . . are employed to reduce difficult mental tasks to simpler ones. Although they are valid in some circumstances, in others they lead to large and persistent biases with serious implications for decision making . . . .").

207. *See, e.g.*, DANIEL KAHNEMAN, THINKING, FAST AND SLOW 7–8 (2011); Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, *in* JUDGMENT AND DECISION MAKING: AN INTERDISCIPLINARY READER 35 (Terry Connolly et al. eds., 2d ed. 2000).

208. *See* Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 COGNITIVE PSYCHOL. 430, 431 (1972) [hereinafter Kahneman & Tversky, *Subjective Probability*] ("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 features of the process by which it is 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 are so many more salesmen than librarians. *See* Tversky & Kahneman, *supra* note 207, at 35.

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 the prior robber's resemblance to 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,[209] it is not necessarily true that a recent robbery was more likely than not 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 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 included[210]—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 is more likely to be 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 additional investigation of *all* suspects, which 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.[211]

Bayes' Theorem helps to formalize the mistake the representativeness bias impels. Adapting one of Kahneman's famous experiments,[212] suppose we know that X, an unidentified bank robber, exhibits behaviors associated

209. *See* LEMPERT ET AL., *supra* note 99, at 483 n.251 (discussing recidivism rates for robbery).

210. *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. L. 7, 42–45 (2010) (discussing susceptibility of police to cognitive biases); *see also* Gudjonsson, *supra* note 106, at 695 (claiming that the lower the base rate of guilt among those interrogated by the police, the greater the risk of a false confession because police interrogate more innocent persons).

211. *See, e.g.*, Ward Casscells et al., *Interpretation by Physicians of Clinical Laboratory Results*, 299 NEW ENG. J. MED. 999, 1000 (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 a particular disease); Itiel E. Dror & David Charlton, *Why Experts Make Errors*, 56 J. FORENSIC IDENTIFICATION 600, 612 (2006) (demonstrating "the vulnerability of experts to contextual effects"); Richard H. Thaler, *Illusions and Mirages in Public Policy*, *in* JUDGMENT AND DECISION MAKING: AN INTERDISCIPLINARY READER, *supra* note 207, at 85, 90 ("Cognitive illusions influence representatives, senators, presidents—even so-called experts are not immune."); Tversky & Kahneman, *supra* note 207, at 50 ("The reliance on heuristics and the prevalence of biases are not restricted to laymen. Experienced researchers are also prone to the same biases . . . .").

212. *See* Tversky & Kahneman, *supra* note 207, at 35.

with the stereotype of a radical bookseller (say, for example, he seems to be 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—multiplied by 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 the culprit is *not* a bookseller.[213] 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. The calculation police detectives might make before deciding to infiltrate or surveil book shops can be represented as follows:

*Prior Odds x Likelihood Ratio = Subsequent Odds*
*(Bs in suspect pool/non-Bs in suspect pool) x (P E(B)) / (P E(not B)) = Subsequent Odds*

Assume that booksellers comprise 5% of the suspect population, so the prior odds are 5 to 95 (1 to 19) that X is a bookseller. Even if 97% of booksellers tend to exhibit the behavior X exhibited, and only 6% of the rest of the population do, there still is a less than even chance (46%) that X is a bookseller:

*Prior Odds x Likelihood Ratio = Subsequent Odds → Odds Stated as Percent Probability*
*1/19 x .97/.06 = .97/1.14 → 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 2%. The probability of guilt is the prior odds that the defendant is guilty *multiplied by* the likelihood ratio associated with the identification. The prior odds are at least partly a function of the number of suspects. If the defendant is one of five suspects, each as likely as the other to have committed the crime, the prior odds of guilt are 1 to 4. If the defendant is one of 5000 such suspects, the prior odds

213. *See supra* notes 60–62 and accompanying text.

are 1 to 4999. 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 3, even though in the former case the probability of guilt is at best only 33% *(1/100 x 1/.02 = 1/2 → 33%)*, while in the latter case the probability of guilt approaches 96% *(1/2 x 1/.02 = 1/.04 → 96%)*.

This last example helps explain why study subjects tend to value eyewitness identifications and confessions more than circumstantial evidence that is as strong or stronger, such as a ballistic match.[214] Unlike an individual who admits guilt or was singled out as the culprit by an eyewitness, 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.[215] 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.[216]

214. *See, e.g.*, Heller, *supra* note 36, 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 LAW & 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) (discussing limitations of subjective evidence); Gary L. Wells, *Naked Statistical Evidence of Liability: Is Subjective Probability Enough?*, 62 J. PERSONALITY & SOC. PSYCHOL. 739, 742 (1992) (similar).

215. In an influential experiment, Professor Wells presented mock jurors with two scenarios in a civil case involving a dog that was run over by a bus. Wells, *supra* note 214, at 741. The bus could only have been operated by either the Blue or Grey Bus Company, and each company had the same number of vehicles. *Id.* 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, a ten-minute drive down the road, but the official acknowledged that his records mistake Grey for Blue Buses 20% of the time. *Id.* In the other case, tire-track marks at the scene of the accident were found to match 80% of the Blue and only 20% of the Grey Buses. *Id.* at 743. Although Wells designed both scenarios to create an equal, 80% probability of Blue Bus liability, judges, psychology students, 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. *Id.*

216. *See* Heller, *supra* note 36, at 244, 255–58 (arguing that juror overvaluation of eyewitness compared to forensic testimony is due not to misevaluation 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 LAW & HUM. BEHAV. 1, 14 (1988) (discussing studies documenting "individuals' reluctance to use statistical information when making causal attributions"); D.H. Kaye & Jonathan J. Koehler, *Can Jurors Understand Probabilistic Evidence?*, 154 J. ROYAL STAT. SOC'Y 75 (1991); Brian C. Smith et al., *Jurors' Use of Probabilistic Evidence*, 20 LAW & HUM. BEHAV. 49, 52–54 (1996) (discussing studies showing persistence of heuristic mistakes even after jurors were trained in

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 especially, as we argue below, on the numerator value of that ratio—and ignore the *prior odds* or effect of the number of potential suspects. When the evidence is overtly probabilistic, however, 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 reasonable doubt that exists, in the common situation in which the prosecution's case is primarily based on an identification or confession and the defendant's evidence is mainly based on a collection of only "small" non-matches.[217] 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 do not 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 do not simply overvalue eyewitness identifications and confessions by ignoring prior odds while giving other evidence of guilt the *correct* weight. Human intuitions systematically give evidence of guilt that is not "direct" or is quite evidently "circumstantial"—including some kinds of forensic evidence—less weight than Bayes' Theorem shows it deserves.[218] 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*."[219]

The representativeness bias helps explain why "direct" evidence is overvalued relative to a correct Bayesian analysis, but it does not explain why "circumstantial" evidence is undervalued. The next three Subparts discuss

Bayesian analysis); Smith et al., *supra* note 43, at 413 (concluding that jurors often overvalue weak and ambiguous forensic evidence). *But see infra* Part VI.A (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).

217. *See supra* notes 1–44 and accompanying text (discussing *Matthews/Hayes* and *DeLuna* cases); *infra* notes 314–20 and accompanying text (discussing *Mayfield* case).

218. *See* Heller, *supra* note 36, at 251–52 (describing a study involving a hypothetical murder case in which five groups of mock jurors were invited to use blood-typing evidence to assess the probability of the defendant's guilt and underestimated the probative value of the evidence relative to the Bayesian norm by 80–100%, 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 L. Smith et al., *supra* note 43, at 410, 414 nn.9–11, 15 (citing sources).

219. Heller, *supra* note 36, at 244–45.

other cognitive biases that explain the latter effect, 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 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.

#### *a. 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.[220] 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.[221]

Heller argues 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.[222] By contrast, "circumstantial" evidence of guilt, even fingerprint or DNA evidence that Bayesian analysis reveals to be stronger, has no simulation effect.[223] It wears a counter-factual on its sleeve: the possibility, however small, that someone else has the same web of lines in a part of his or her fingerprint, or that a different firearm was used that leaves the same

220. *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. EXPERIMENTAL SOC. PSYCHOL. 88, 90–92 (1978) (describing research indicating that imagining a scenario increases subjects' assessment of its likelihood); Heller, *supra* note 36, at 260–61; Daniel Kahneman & Amos Tversky, *The Simulation Heuristic*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, *supra* note 206, 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 when experiencing an actual event makes it seem more foreseeable than it was. *See, e.g.*, Hartmut Blank et al., *Hindsight Bias: On Being Wise After the Event*, 25 SOC. COGNITION 1 (2007).

221. *See* Heller, *supra* note 36, at 260–61.

222. *See id.* at 296. "Coverage" refers to the proportion of the overall event that the evidence portrays or "covers."

223. *See id.* at 265.

pattern of striations on bullets it fires as the firearm in evidence in the case, or that the defendant was framed. Adding to the disparity, direct evidence in the form of an eyewitness identification or confession is "vivid," meaning that it is "representational" or seems to present a single reality; "narrative" in that it comes in the especially accessible form of a relatively coherent story; "univocal" in that it points in a single direction; and appears to be "unconditional."[224] By contrast, circumstantial evidence is "pallid," "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."[225]

The vividly simulating effect of eyewitness statements and confessions is likely to sway police and prosecutors when they decide 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 almost everything that happened, they will "tend to cease the simulation process and fail to consider alternative scenarios that may imply a different outcome."[226] The effect may be somewhat mitigated at trial because defense counsel during trial and jurors during deliberations 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.[227] Insofar, however, 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 or get any of the extra "simulation" or "representativeness" bounce that eyewitness and confession evidence gets.[228] 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.[229]

---

224. *See id.* at 264–69.

225. *See id.*; *see also* Richard K. Greenstein, *Determining Facts: The Myth of Direct Evidence*, 45 HOUS. L. REV. 1801, 1815 (2009) (offering a 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, which is connected to the fact of consequence [only] by inferential steps").

226. *See* Edward R. Hirt & Keith D. Markman, *Multiple Explanation: A Consider-an-Alternative Strategy for Debiasing Judgments*, 69 J. PERSONALITY & SOC. PSYCHOL. 1069, 1084 (1995).

227. *See* Heller, *supra* note 36, 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 whether the nature of the decision itself involves 'negative affect' such as fear or regret").

228. *See id.* at 292–96 (arguing that reasonable-doubt and circumstantial-evidence defenses deprive jurors of the sorts of evidence that are most capable of combatting confirmation bias).

229. *See id.* at 285–88. Heller describes experimental evidence of "truth bias," a tendency to believe another's autobiographical statements regardless of their truthfulness, *id.* at 285–86,

### *b. 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.[230] Professor 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.[231]

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 who she initially

and "narrative transportation," which leads listeners to accord more weight to the credibility of statements in the form of stories. *Id.* at 287–88. Because of truth bias, for example, 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. *Id.* at 285; *see also* Melanie C. Green & Timothy C. Brock, *The Role of Transportation in the Persuasiveness of Public 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 36, at 287–88 (presenting evidence indicating that 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").

230. *See, e.g.*, KAHNEMAN, *supra* note 207, at 209–12 (discussing the illusion of validity); Karl Ask & Pär Anders Granhag, *Motivational Sources of Confirmation Bias in Criminal Investigations: The Need for Cognitive Closure*, 2 J. INVESTIGATIVE PSYCHOL. & 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) [hereinafter Burke, *Improving Prosecutorial Decision Making*]; Alafair Burke, *Neutralizing Cognitive Bias: An Invitation to Prosecutors*, 2 N.Y.U. J.L. & LIBERTY 512, 516–20 (2007); Keith A. Findley & Michael S. Scott, *The Multiple Dimensions of Tunnel Vision in Criminal Cases*, 2006 WIS. L. REV. 291, 309–17; Gudjonsson, *supra* note 106, at 699; Charles G. Lord et al., *Considering the Opposite: A Corrective Strategy for Social Judgment*, 47 J. PERSONALITY & SOC. PSYCHOL. 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); Paul C. Giannelli, *Confirmation Bias*, CRIM. JUST., Fall 2007, at 60, 61.

231. *See* Barbara O'Brien, *Prime Suspect: An Examination of Factors That Aggravate and Counteract Confirmation Bias in Criminal Investigations*, 15 PSYCHOL. PUB. POL'Y & L. 315, 328–29 (2009).

identified.[232] Once the representativeness and simulation biases lead investigators or jurors to overvalue an eyewitness's or a confessing suspect's narrative about what happened, those individuals will tend to "adhere to their beliefs [though] the original evidential basis of the beliefs is shown to be flimsy, false, or nonexistent,"[233] and this obeisance will likely deter them from searching for alternative theories even after finding out that the witness may well be mistaken or untruthful.[234] The perseverance of beliefs based on visibly discredited evidence gets even stronger when the discredited information is part of a "coherent, causally related account in which a single or minimal correction has a significant impact on the construal of meaning."[235] The bias is stronger, therefore, 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.[236]

Notice that every attribute of "circumstantial" evidence that keeps it from getting the artificial representativeness, simulation, and confirmation bounces even more clearly diminishes the weight decision makers are likely to accord to the non-exclusionary non-match evidence of innocence that interests us here. Non-exclusionary non-matches are as pallid, abstract, rhetorical, non-transporting, polyvocal, and probabilistic as evidence can be.[237] They are non-representational and non-narrative; often fall in a

---

232. *See id.* at 318; Joep Sonnemans & Frans van Dijk, *Errors in Judicial Decisions: Experimental Results*, 28 J.L. ECON. & ORG. 687, 712–13 (2012) (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).

233. Martin F. Davies, *Belief Persistence After Evidential Discrediting: The Impact of Generated Versus Provided Explanations on the Likelihood of Discredited Outcomes*, 33 J. EXPERIMENTAL SOC. PSYCHOL. 561, 562 (1997); *see also* Craig A. Anderson, *Abstract and Concrete Data in the Perseverance of Social Theories: When Weak Data Lead to Unshakeable Beliefs*, 19 J. EXPERIMENTAL SOC. PSYCHOL. 93, 95 (1983).

234. *See* Heller, *supra* note 36, 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." (emphasis omitted)); 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").

235. Heller, *supra* note 36, at 289 (quoting Hollyn M. Johnson & Colleen M. Seifert, *Sources of the Continued Influence Effect: When Misinformation in Memory Affects Later Influences*, 20 J. EXPERIMENTAL PSYCHOL.: LEARNING MEMORY & COGNITION 1420, 1432 (1994)) (internal quotation marks omitted).

236. *Id.* at 290 ("[B]elief-perseverance is strongest when the evidence supporting a belief is concrete and causally coherent. Circumstantial evidence is neither . . . ."); *see also* Sykes & Johnson, *supra* note 214, 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 "a belief engendered by an assertion about an event is more difficult to mutate than a belief based solely on statistical probabilities").

237. *See supra* notes 223–29.

demeaned category of the "absence" or "negation" of evidence;[238] typically qualify as "small" evidence given their limited probative weight; by definition 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 convincing counter-theories and thus carry the seeds of its own destruction is particularly true of *non-exclusionary non-match evidence* of *innocence*.[239]

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 by the likelihood ratio, with the latter defined as the probability that the evidence would exist if the suspect is guilty divided by the probability that 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).

#### *c. The Certainty Effect*

Another bias, the "certainty effect,"[240] helps explain why lay decision makers give less weight to circumstantial evidence 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 tend to "overweight outcomes that are considered certain, relative to outcomes

238. *See* Stephen A. Saltzburg, *A Special Aspect of Relevance: Countering Negative Inferences Associated with the Absence of Evidence*, 66 CALIF. L. REV. 1011, 1019–21 (1978) (criticizing the use of negative inferences from the absence of evidence in making factual determinations in court cases).

239. Heller, *supra* note 36, at 299–300. Professor Heller stated that "jurors generally find it relatively easy to imagine [counter-theories] in a circumstantial case. The polyvocity of circumstantial evidence means that the prosecution's own evidence is available for use in [counter-theories], and strong priming normally ensures that jurors will pay close attention to any" competing evidence and counter-theory offered. *Id.* at 300.

240. *See* KAHNEMAN, *supra* note 207, at 311–12; Daniel Kahneman & Amos Tversky, *Prospect Theory: An Analysis of Decision Under Risk*, *in* CHOICES, VALUES, AND FRAMES 17, 19–22 (Daniel Kahneman & Amos Tversky eds., 2000).

which are merely probable"[241] and "greatly undervalue a reduction in the probability of a hazard in comparison to the complete elimination of that hazard."[242]

There is a built-in fudge factor, as well. Naïve decision makers tend to treat the *nearly* complete reduction of even a very serious risk as if it removed *all* risk, creating a false sense of security.[243] "[S]tudies of insurance markets have shown that we tend to ignore small risks until their probability passes a certain threshold, at which point we overspend wildly to prevent them."[244] 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 it[245]—or an excessively worrisome probability of harm. Subjects asked to say whether they would volunteer to receive a vaccine that halves the risk of contracting a serious disease expected to afflict 20% of the population are substantially less likely to volunteer than subjects invited to receive a vaccine that would reduce to *zero* the probability of contracting one of two equally serious strains of the disease but have no effect on the other strain when each strain afflicts 10% of the population.[246] Although the risk reduction in both cases is the same—from 20 to 10%—the description of a treatment as reducing one of two equal risks to zero evidently 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 the evidence as establishing a "certainty" of guilt than even very strong circumstantial evidence that creates a much higher probability of guilt.[247] Jurors may reach a certainty conclusion about eyewitness and

241. Kahneman & Tversky, *supra* note 240, at 20.

242. Daniel Kahneman & Amos Tversky, *Choices, Values, and Frames*, *in* CHOICES, VALUES, AND FRAMES, *supra* note 240, at 9; *see also* Kuran & Sunstein, *supra* note 206, 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., *Risk as Feelings*, 127 PSYCHOL. BULL. 267, 276 (2001) (showing that individuals will pay far more to reduce their risk of poisoning "from 5 in 10,000 to 0 than from 15 in 10,000 to 5 in 10,000," though the latter reduction is twice as large); Thaler, *supra* note 211, at 87–88 (similar).

243. *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 . . . .").

244. Benjamin Wallace-Wells, *Cass Sunstein Wants To Nudge Us*, N.Y. TIMES MAG. (May 13, 2010), http://www.nytimes.com/2010/05/16/magazine/16Sunstein-t.html?_r=2&pagewanted=all (discussing RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008)).

245. Amos Tversky & Daniel Kahneman, *The Framing of Decisions and the Psychology of Choice*, 211 SCIENCE 453, 455 (1981).

246. Paul Slovic et al., *supra* note 206, at 480–81 (discussing vaccination study).

247. Heller, *supra* note 36, at 283–85.

confession evidence, Heller suggests, because such evidence establishes a 100% 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 is wrong gives jurors more confidence in the accuracy of the evidence than it deserves.[248] By contrast, circumstantial evidence of guilt—no matter how reliable the testimony presenting it—always comes with some doubt.[249] That doubt is magnified by the stress jurors experience when contemplating the possibility of an inaccurate verdict.[250] As a result, Heller reasons, "jurors 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.[251]

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 . . . . [s]ince 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.[252] If certainty arises only in the perceived absence of *any* overt possibility that the hazard will generate a harm,[253] there is no reason why a 1% chance that a witness is lying should trigger any less anxiety than a 1%, or even 0.0000001% 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 are mistaken or lying, or that a confession is untrue, 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 "[t]he certainty effect says that jurors 'overweight outcomes that are *considered* certain,' not outcomes that *are* certain."[254] 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 heterodox perceptions

248. *See id.* at 268; *supra* notes 220–21, 233–36 and accompanying text.

249. *See* Heller, *supra* note 36, at 268.

250. *See id.* at 282 ("Jurors believe that they 'should make accurate determinations with respect to the actual guilt or innocence of the defendant,' and they experience considerable stress in trying to comply with that self-imposed mandate. Moreover, jurors consistently report that choosing a verdict is the most stressful aspect of a criminal case." (quoting William C. Thompson et al., *Inadmissible Evidence and Juror Verdicts*, 40 J. PERSONALITY & SOC. PSYCHOL. 453, 454 (1981) (footnote omitted))).

251. *Id.* at 284.

252. *Id.* at 284–85 (internal quotation marks omitted).

253. *See supra* notes 240–46 and accompanying text.

254. Heller, *supra* note 36, at 285 (quoting Kahneman & Tversky, *supra* note 240, at 20).

of a dozen demographically diverse jurors[255]—designed to rub jurors' noses in facts of uncertainty.

Part of the problem, we believe, lies with Heller's distinction between "direct" evidence (primarily eyewitness testimony and confessions) and "circumstantial" evidence (such as 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 although 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 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 Subpart, 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 Subpart, we hypothesize the opposite craving, namely 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 in fact it does not.[256] 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

255. *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.").

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.[257]

If the "certainty" achieved by reducing risk to zero creates vastly more advertent pleasure for people than other comparable or greater reductions of risk,[258] 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 does not 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 may be important psychological benefits to believing in the truth of that equation, triggering the "certainty" bounce even in the face of an obviously more accurate judgment that may generate 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 possibility of making a mistake that leaves a killer at large or convicts an innocent person, giving them every psychological incentive to find shelter in even a false sense of certainty.[259]

The Bayesian equation (*prior odds* × *likelihood ratio* = *subsequent odds*) again helps model the heuristic process we describe. The values in the likelihood ratio (the probability that evidence would exist 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 an intuitive and easily calculated measure of the probative weight of evidence through another equation:

*(numerator - denominator) / numerator = probative value.*[260]

256. *See supra* note 245 and accompanying text.

257. JAMES BOSWELL, THE LIFE OF SAMUEL JOHNSON 301 (David Womersley ed., Penguin Books 2008) (1791).

258. *See supra* notes 240–46 and accompanying text.

259. *See supra* note 250 and accompanying text.

260. *See* Lempert, *supra* note 53, at 1047. As we note above, it is conventional to define probative weight as the likelihood ratio itself (i.e., its numerator *divided by* its denominator). *See supra* note 53 and accompanying text. We prefer here, however, to use a different definition—numerator minus denominator—to provide a simple reflection of the fact that probative weight is based on a comparison of the strength of the numerator and the denominator probabilities: if they are equal, there is no probative weight; if they are very different in size, probative weight is high. Subtracting one from the other thus nicely illustrates probative value in a computationally simple way: if Judge *A* evaluates a piece of evidence by estimating a numerator value of .7 and a denominator value of .2, and if Judge *B* provides a different estimate of, say, .5 for the denominator and .4 for the numerator, it is easy to see that Judge *A* values the evidence five times more (*.7 - .2 = .5*) than Judge *B* (*.5 - .4 = .1*). When, however, the comparison isn't between two different assessments of the *same* evidence but between two assessments of *different*

Notice that a high *numerator* is *necessary* to a very high probative value but is not sufficient for it. Very weighty evidence requires a low denominator as well. This construct suggests a shortcut that 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 demonstration that human decision makers seem to anchor on large numerator probabilities to the exclusion of the denominator is Kahneman's famous "Linda" experiment.[261] Test subjects read a profile of "Linda," who was a "31 years old, single, outspoken and very bright" woman who was a philosophy major in college, worried about "discrimination and social justice" and "participated in anti-nuclear demonstrations." [262] Subjects were then asked to predict the most factually likely of two descriptions of Linda: that "Linda is a bank teller and is active in the feminist movement" or that "Linda is a bank teller."[263] Most subjects committed the logical fallacy of believing that the former probability is greater than the latter, though the category of bank tellers is perforce larger because it contains the category of

pieces of evidence, it is necessary to use the more complicated equation in text to provide a common scale for comparison: subtracting the denominator from numerator then dividing by the numerator. More precisely, we should say that probative weight is the *absolute value* of that result of that equation, reflecting the fact that a numerator of .7 and a denominator of .2 yields evidence that is equally weighty as evidence with a numerator of .2 and a denominator of .7. The former evidence supports a hypothesis of guilt, and the latter supports a hypothesis of innocence, but the evidence in each case is equally strong.

261. Amos Tversky & Daniel Kahneman, *Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment*, 90 PSYCHOL. REV. 293, 297 (1983).

262. *Id.*

263. *Id.* at 299.

bank tellers active in the feminist movement.[264] 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 people 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.'"[265] 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.

A study by Beyth-Marom and Fischhoff provides further evidence that decision makers pay closer attention to the numerator than the denominator probability.[266] They asked subjects to identify information that 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.[267] Two of the pieces of information that subjects could request were the percentage of professors at the party who were members of the Bear's Club (in Bayesian terms, the numerator probability) and the percentage of business executives at the party who were members of the Bear's Club (the denominator probability).[268] Most subjects wanted to know the former probability; few cared about the latter one.[269]

The "uniqueness fallacy" also appears to be a prevalent and longstanding concern in everyday trials. Consider Federal Rules of Evidence 404 through 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.[270] 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, as when someone follows-up an accident with a so-called

264. *Id.* at 299–300; *see also* KAHNEMAN, *supra* note 207, at 328–29 (discussing "denominator neglect").

265. STEPHEN JAY GOULD, BULLY FOR BRONTOSAURUS: REFLECTIONS IN NATURAL HISTORY 469 (1991).

266. Ruth Beyth-Marom & Baruch Fischhoff, *Diagnosticity and Pseudodiagnosticity*, 45 J. PERSONALITY & SOC. PSYCHOL. 1185 (1983).

267. *Id.* at 1186.

268. *Id.* at 1187.

269. *Id.* at 1188.

270. FED. R. EVID. 404(a)(1), 404(b)(1).

"subsequent remedial measure,"[271] offers to settle a civil claim,[272] pays the medical expenses of an accident victim,[273] cops a plea to a crime,[274] or insures himself against liability for accidents.[275] In the same category are evidence doctrines discouraging inferences of a consciousness of guilt from silence in the face of a criminal accusation,[276] from a suspect's flight from arresting police officers,[277] or from a refusal to take a polygraph test.[278]

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 act guilty are guilty, without considering innocent explanations for the behavior.[279] Rephrased in Bayesian terms, the law fears that 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 probability without considering a non-inconsequential denominator probability. The law consequently excludes the evidence to be sure that jurors do not treat it as *unique* to guilty people (i.e., 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 in the face of accusations, the law expects jurors to assume that anyone who has done one of these things is guilty and ignore the fact that innocent people often do them too. Careful people, that is, may quickly repair unanticipated hazards, and innocent people may worry that polygraphs will mistake nervousness for guilt, but the law expects the high numerator value to keep jurors from considering these possibilities.

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.[280] The forbidden evidence either reveals a trait resembling that of a guilty person (prior bad acts) or

271. *Id.* 407.

272. *Id.* 408(a)(2).

273. *Id.* 409.

274. *Id.* 410(a)(1).

275. *Id.* 411.

276. *See, e.g.*, Doyle v. Ohio, 426 U.S. 610, 619 (1976) (White, J., concurring) (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).

277. *See, e.g.*, KENNETH S. BROUN ET AL., MCCORMICK ON EVIDENCE 406–07 (6th ed. 2006).

278. *See, e.g.*, deVries v. St. Paul Fire & Marine Ins. Co., 716 F.2d 939, 945 (1st Cir. 1983) (holding inadmissible a witness's refusal to take a polygraph exam); Aetna Ins. Co. v. Barnett Bros., Inc., 289 F.2d 30, 34 (8th Cir. 1961) (same). *See generally* BROUN et al., *supra* note 277, at 403–13 (noting the evidence rules' resistance to inferences of a consciousness of guilt from conduct).

279. *See, e.g.*, LEMPERT ET AL., *supra* note 99, at 279–81, 333–41 (discussing that evidence rules are designed to counteract jurors' tendency to overestimate the value of evidence of conduct in which guilty people often engage).

280. *See supra* Part V.A.2.

simulates the endgame of many crimes (actions to avoid apprehension). Second, contrary to what Heller predicts,[281] every one of these examples is triggered by "circumstantial"—rather than "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 are not. Contrary to Heller's assumption, however, they are "circumstantial" not *only* because they depend upon probabilistic inferences of witness credibility[282] but also because they depend upon aggregations of many only modestly "probable" individual matches.[283] 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 dog barked, then a light was turned on) 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 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.[284] Recent scholarship also criticizes the myth of "uniqueness" and "individualization" as to tool marks, handwriting, bite marks, shoe prints, and fingerprints.[285] Fingerprints are especially

281. *See supra* text accompanying notes 247–51.

282. *See* Heller, *supra* note 36, at 247 (assuming that credible eyewitness identifications and confessions have a probative value of 1, or 100%).

283. *See supra* Part IV.A.1.

284. *See supra* notes 117–21 and accompanying text; *see also* LEVY, *supra* note 113, at 26 (noting the misimpression of DNA as "a genetic fingerprint" that "is individually specific" and "does not belong to any other present or future person on earth"); Jonathan J. Koehler, *When Are People Persuaded by DNA Match Statistics?*, 25 LAW & 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 essentially zero); Smith et al., *supra* note 43, 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); Neufeld & Colman, *supra* note 94, at 50 (decrying the misimpression that DNA "identifies the 'genetic code' unique to an individual and indeed is as unique as a fingerprint").

285. *See, e.g.*, Simon A. Cole et al., *Beyond the Individuality of Fingerprints: A Measure of Simulated Computer Latent Print Source Attribution Accuracy*, 7 LAW PROBABILITY & RISK 165, 165

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.[286] 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 that does not invoke that bias, such as drops of blood and mazes of lines left by oily human hands. 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," which disposes naïve decision makers to perceive uniqueness where none exists, is an "irrelevance fallacy," which 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 weight subjects give to a match between a rare blood-type and hair characteristics of an unknown perpetrator and an identified defendant.[287] The authors were "most surpris[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."[288] An example of this fallacy is the statement that "even though only 2% of the population has characteristic X, in the entire population of this city, there are hundreds of people with that rare trait."[289] This is a fallacy because the realistic number of suspects in most cases is smaller than the entire population, so that it remains true that any member of the suspect group is unlikely to have 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

(2008); Cole, *supra* note 102, at 234–36, 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"); Jonathan J. Koehler & Michael J. Saks, *Individualization Claims in Forensic Science: Still Unwarranted*, 75 BROOK. 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 the "individualization fallacy"); sources cited *supra* note 96.

286. *Definitions for Fingerprint*, DEFINITIONS, http://www.definitions.net/definition/Fingerprint (last visited Oct. 11, 2012) (providing one of Random House Webster's College Dictionary's definitions of fingerprint as "any unique or distinctive pattern that presents unambiguous evidence of a specific person, substance, disease, etc."). *But see supra* note 96 and accompanying text (explaining why fingerprint evidence is not unique).

287. *See* Thompson & Schumann, *supra* note 167, at 182 n.6.

288. *Id.* at 182.

289. *See id.* at 171 (offering a similar example).

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 similar to the bias the "Linda" study reveals.[290] The evidence is weak, to be sure, because the difference between the numerator and denominator is small. But the irrelevance fallacy predicts that the jury will assume 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 arguing to the jury that the affray must have dislodged the cigarette from the mouth of a smoker–perpetrator who is not the defendant, the prosecutor can easily counter with far more likely scenarios in which a non-perpetrator dropped the butt.

Now, assume there are two such non-matches—the cigarette butt and a stray maroon button at the scene that is not associated with the victim's clothing or with the defendant's clothing when he was arrested right after the crime. As we develop above, although both non-matches are weak evidence of innocence, 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 attention, there are only zeros to aggregate.

Recent studies by Lisa L. Smith and colleagues are also of interest here. They 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.[291] Each subject considered several 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 of 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).[292] Based on these features, the authors created four categories of evidence (high relevance/low

290. *See supra* notes 261–64 and accompanying text.

291. Smith et al., *supra* note 43, at 410.

292. *Id.* at 410–11.

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."[293] Evincing considerable facility with circumstantial evidence, subjects drew sharp and appropriate distinctions among different items of evidence, recognizing that some were strong, some were weak, and some fell in between.[294]

In a second study, the researchers presented three groups of subjects with descriptions of evidence at a murder trial.[295] Each group received the same evidence (mainly witness statements) with one exception.[296] One group was told that strong DNA evidence (high relevance/low mobility) linking the defendant to the crime was also present.[297] Another group was told that moderately probative DNA evidence (high relevance/high mobility) was present. A third group was told that weak DNA evidence (low relevance/high mobility) was present.[298] The mock jurors estimated the strength of the evidence twice: after hearing all the evidence and after hearing closing statements.[299] At the latter stage, the subjects also rendered a verdict.[300] 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 guilt in the *weak* DNA case.[301] So far the news is all of the dog-bites-man variety.[302]

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%), even though their estimates of the probability of guilt were only slightly different in the two cases.[303] 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.[304] 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

293. *Id.*
294. *See id.* at 411.
295. *Id.* at 411–12.
296. *Id.* at 412.
297. *Id.*
298. *Id.*
299. *Id.*
300. *Id.*
301. *Id.* at 412–13.
302. *See supra* notes 249–51 and accompanying text.
303. Smith et al., *supra* note 43, at 413.
304. *Id.*

"uniqueness" conclusion (proving guilt) or could identify alternative explanations for the DNA evidence and ignore it (leading to acquittal).[305] This may provide evidence of the uniqueness fallacy in the former case and the irrelevance fallacy in the latter one.

A final set of experimental studies suggests that we have understated the tendency of decision makers to misevaluate the importance of "small" evidence, such as non-exclusionary non-matches. Rather than treating weak evidence of a factual proposition as having *no* value, lay decision makers seem to treat it as having a *negative* value (i.e., as enhancing confidence in the contrary proposition). These studies also support a hypothesis introduced above, that the irrelevance fallacy increases the impact of the uniqueness fallacy.

Each of the studies in question presents subjects with two scenarios, one involving only strong evidence, say, of a defendant's guilt, *plus* weak evidence that either supports the same proposition or supports a contrary conclusion. In each study, subjects gave the weak evidence the opposite effect that Bayesian analysis requires. When the strong and weak evidence both supported the same conclusion, the inclusion of the weak with the strong evidence made decision makers *less* likely to accept the proposition than when only the strong evidence was presented. When the strong evidence supported one proposition and the weak evidence supported a contrary conclusion, the inclusion of the weak evidence made decision makers substantially *more* likely to reach, and more confident in, the conclusion supported by the strong evidence then when the strong evidence was considered by itself.[306] Of course, if decision makers were proceeding in

305. *See id.* at 412–13; *see also* Sonnemans & van Dijk, *supra* note 232, at 27 (reporting research conclusions 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 eyewitness identifications because they appear to match a unique trait of the defendant to that of the perpetrator but underappreciate an eyewitness's inability to identify the defendant. *See supra* notes 251–53 and accompanying text; *see also* Hunter A. McAllister & Norman J. Bregman, *Juror Underutilization of Eyewitness Nonidentifications: Theoretical and Practical Implications*, 71 J. APPLIED PSYCHOL. 168, 169–70 (1986).

306. *See, e.g.*, Philip M. Fernbach et al., *When Good Evidence Goes Bad: The Weak Evidence Effect in Judgment and Decision-Making*, 119 COGNITION 459, 460, 462–63, 465 (2011) (providing experimental evidence of the "weak evidence effect," and concluding that the presentation of "weak positive evidence . . . could actually be deleterious"); James Friedrich & Paul Smith, *Suppressive Influence of Weak Arguments in Mixed-Quality Messages: An Exploration of Mechanisms via Argument Rating, Pretesting, and Order Effects*, 20 BASIC & APPLIED SOC. PSYCHOL. 293, 293–94, 298–302 (1998) (citing numerous studies as well as their own findings that "[w]eak arguments were not simply ignored or discarded; they actually undermined the persuasiveness of the message as a whole"); Adam Harris et al., *"Damned by Faint Praise": A Bayesian Account*, *in* PROCEEDINGS OF THE 31ST ANNUAL CONFERENCE OF THE COGNITIVE SCIENCE SOCIETY 292, 292 (N.A. Taatgen & H. van Rijn eds., 2009), *available at* http://www.ucl.ac.uk/lagnado-lab/publications/harris/FPconference.pdf (citing studies supporting the "Boomerang effect" of weak evidence); Lola L. Lopes, *Procedural Debiasing*, 64 ACTA PSYCHOLOGICA 167, 170 (1987)

the additive fashion that Bayesian logic requires, the result would be the opposite: strong evidence bolstered by weak evidence would be slightly more persuasive than the stronger evidence by itself, and strong evidence diminished by weak contradicting evidence would be slightly less likely to convince than the strong evidence by itself.[307] The studies further reveal that (1) the more weak evidence that is presented, the greater the "boomerang" effect it has;[308] (2) the bias arises even when both the strong and weak evidence are obviously "circumstantial";[309] and (3) the bias appears even when experimental conditions rule out the possibility that subjects treated the weakness of the evidence of X as proof that X probably was not true, as might occur if subjects assumed that stronger evidence of X would have been presented if X were true.[310]

This so-called "weak evidence effect" appears to result from an increase in confidence in the proposition supported by the strong evidence when its power is compared to that of the weak evidence.[311] Decision makers, that is, seem to act as if each new item of weak evidence supporting the same proposition as the strong evidence is averaged into the numerator value in the Bayesian likelihood ratio, causing probative value to diminish,[312] while each weak item supporting a proposition contrary to the strong one is treated as representing the denominator value, causing the estimated probative value of the evidence to increase because the difference between

(citing studies); Craig R.M. McKenzie et al., *When Negative Evidence Increases Confidence: Change in Belief After Hearing Two Sides of a Dispute*, 15 J. BEHAV. DECISION MAKING 1, 2, 4, 7–8, 13 (2002) (discussing and presenting studies demonstrating the "reverse impact" of weak evidence).

307. *See* Harris et al., *supra* note 306, at 292 ("[A]ssuming the evidence is not known to be misleading, evidence in favor of the hypothesis, no matter how weak, can never decrease the person's degree of belief in the hypothesis.").

308. *See, e.g.*, Friedrich & Smith, *supra* note 306, at 294 (citing studies finding that belief in the proposition supported by weak evidence "did, in fact, diminish as increasing numbers of weak arguments were added").

309. *See, e.g.*, McKenzie et al., *supra* note 306, at 6–8 (using examples involving evidence that is conventionally understood as direct and evidence conventionally understood as circumstantial).

310. *See, e.g.*, Fernbach et al., *supra* note 306, at 462 ("The results are inconsistent with the pragmatic explanation that weak evidence is interpreted as negative with respect to a default expectation [of stronger evidence than was presented]."); Friedrich & Smith, *supra* note 306, at 298–300; McKenzie et al., *supra* note 306, at 6–8, 14 (noting questions put to subjects indicating that they understood the weak evidence as (weakly) suggesting innocence, even though its introduction made them more likely to convict than when the weak evidence was not presented); *cf.* Harris et al., *supra* note 306, at 292, 296 (identifying circumstances in which the reverse effect of weak evidence is consistent with a reasonable inference that the evidence presented would not be weak unless the opposite proposition were true under the circumstances).

311. *See, e.g.*, McKenzie et al., *supra* note 306, at 2–13.

312. *See, e.g.*, Lopes, *supra* note 306, at 170 (discussing studies in which subjects appear to be engaging in such averaging).

the numerator and denominator seems to have increased.[313] Weak contrary evidence thus appears to operate by magnifying the "uniqueness" effect of "big" evidence—causing the evidence to appear stronger—even as the irrelevance fallacy leads the decision maker to ignore "small" evidence entirely.

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 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.[314] 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.[315]

The FBI based Mayfield's arrest on a "100% match" between his fingerprint and partial prints that Spanish authorities had found on a bag of detonating devices shortly after the explosions.[316] 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.[317]

313. *See, e.g.*, McKenzie et al., *supra* note 306, at 4–7, 14 (presenting findings suggesting that the extent of the reverse effect of weak evidence increases as the difference in probative value between the strong and weak evidence increases).

314. *See* STEVEN T. WAX, KAFKA COMES TO AMERICA: FIGHTING FOR JUSTICE IN THE WAR ON TERROR 6 (2008).

315. *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 . . . ."), *vacated*, 599 F.3d 964 (9th Cir. 2010); Susan Schmidt, *American Held in Madrid Bombings*, DAILY (University of Washington) (May 7, 2004), http://dailyuw.com/archive/2004/05/07/imported/american-held-madrid-bombings#.UJ6TGYIe4W8.

316. *Mayfield*, 504 F. Supp. 2d at 1029. "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." Press Release, FBI, Statement on Brandon Mayfield Case (May 24, 2004), *available at* http://www.fbi.gov/news/pressrel/press-releases/statement-on-brandon-mayfield-case; *see also* Steven T. Wax & Christopher J. Schatz, *A Multitude of Errors: The Brandon Mayfield Case*, 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 . . . Brandon Mayfield.").

317. OFFICE OF THE INSPECTOR GEN. OVERSIGHT & REVIEW DIV., U.S. DEP'T OF JUSTICE, A REVIEW OF THE FBI'S HANDLING OF THE BRANDON MAYFIELD CASE 82–83 (2006), *available at* http://www.justice.gov/oig/special/s0601/final.pdf.

Mayfield's 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 ranked fourth on the list,[318] perhaps in part because he was married to an Egyptian woman, converted to Islam, and had provided legal services in a child-custody case to a man sentenced to prison for attempting to travel to Afghanistan to join the Taliban.[319] The FBI subsequently 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.[320]

Mayfield's case was rife with heuristic traps. The representativeness and simulation biases associated with Mayfield's conversion to Islam and representation of a man with pro-Taliban sympathies evidently triggered a powerful, if imaginary, scenario of a turn towards Islamic terrorism that blinded officials to the other suspects in the FBI database with "similar" fingerprints, not to mention additional suspects that a simple extrapolation of that number to international databases would have suggested. Via the confirmation bias, the "100% match" and accompanying scenario 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, including the justice system, use or could use to head off debilitating biases.[321] Above, we note how Federal

318. *See Mayfield*, 504 F. Supp. 2d at 1027.

319. *See* WAX, *supra* note 314, at 6–7.

320. *See* Dan Eggen, *U.S. Settles Suit Filed by Ore. Lawyer*, WASH. POST (Nov. 30, 2006), http://www.washingtonpost.com/wp-dyn/content/article/2006/11/29/AR2006112901179.html; *see also* Mayfield v. United States, 599 F.3d 964, 968 (9th Cir. 2010) (describing FBI's agreement to pay compensatory damages of $2 million, destroy the fruits of its surveillance, and apologize); *Apology Note*, WASH. POST (Nov. 29, 2006), http://www.washingtonpost.com/wp-dyn/content/article/2006/11/29/AR2006112901155.html.

321. *See, e.g.*, COMM. ON BEHAVIORAL & SOC. SCI. RESEARCH TO IMPROVE INTELLIGENCE ANALYSIS FOR NAT'L SEC. ET AL., NAT'L RESEARCH COUNCIL, INTELLIGENCE ANALYSIS FOR TOMORROW: ADVANCES FROM THE BEHAVIORAL & SOCIAL SCIENCES 84–87 (2011) (recommending various ways of quantifying and aggregating probabilities to increase the rigor and performance evaluation of intelligence analysis); KAHNEMAN, *supra* note 207, 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 useful checklists, as well as more elaborate exercises, such as reference-class forecasting and the premortem."); THALER & SUNSTEIN, *supra* note 244, at 177–78; Burke, *Improving Prosecutorial Decision Making*, *supra* note 230, 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 Vepsalainen, *The Economic Efficiency of Corporate Culture*, 9 STRATEGIC

Rules of Evidence 404 through 411 diminish the representativeness bias and tendency to overvalue the numerator and undervalue the denominator in assessing evidence.[322] Likewise, "devil's advocate" mechanisms can mitigate simulation, certainty, and confirmation bias.[323] Such mechanisms force people to imagine scenarios and present arguments that counter their initial dispositions,[324] or assign decisions to large and diverse groups likely to generate competing scenarios on their own.[325] This literature raises two questions. First, do our existing legal procedures and rules sufficiently

MGMT. J. (SPECIAL ISSUE) 115 (1988); Lynne L. Dallas, *The New Managerialism and Diversity on Corporate Boards of Directors*, 76 TUL. L. REV. 1363, 1407 (2002) (discussed *infra* note 325); Kenneth R. 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, 209–12 (2006) (arguing for legal policies that help eliminate problems generated by "bounded rationality" and cognitive biases); Lopes, *supra* note 306, at 173, 178–79 (presenting evidence that the weak-evidence effect can be diminished by causing decision makers expressly to acknowledge the value (however weak) and the direction of the inferences supported by weak evidence before considering it along with other evidence); Jesse M. Pines, *Profiles in Patient Safety: Confirmation Bias in Emergency Medicine*, 13 ACAD. EMERGENCY MED. 90, 92–93 (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 LAW & SOC'Y REV. 123, 131 (1980) (describing many areas in which decision aids are utilized to prevent bias).

322. *See supra* notes 270–75 and accompanying text.

323. The name, of course, comes from mechanisms that the Catholic Church has long used to increase the reliability of canonization decisions. THE CATHOLIC ENCYCLOPEDIA (Charles G. Herbermann et al. eds., 1907) (stating that the duty of "Advocatus Diaboli" is "to prepare in writing all possible arguments, even at times seemingly slight, against the raising of any one to the honours of the alter"); *see also, e.g.*, Craig A. Anderson & James J. Lindsay, *The Development, Perseverance, and Change of Naive Theories*, 16 SOC. COGNITION 8, 24 (1998) (using a "counter-explanation" process in which the subject "imagines and explains how a different [scenario] is (or might be) true" to counteract confirmation bias); Burke, *Improving Prosecutorial Decision Making*, *supra* note 230, 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 ORGANIZATIONAL BEHAV. & HUM. DECISION PROCESSES 135, 136 (1997); Lord et al., *supra* note 230, at 1238–40 (finding that telling experimental subjects to consider the opposite of their initial hypothesis induced more accurate results than instructing them to be fair and unbiased).

324. *See, e.g.*, Findley & Scott, *supra* note 230, at 371; Adam D. Galinsky & Gordon B. Moskowitz, *Counterfactuals as Behavioral Primes: Priming the Simulation Heuristic and Consideration of Alternatives*, 36 J. EXPERIMENTAL SOC. PSYCHOL. 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).

325. The Constitution arguably adopts this strategy in criminal cases by requiring that juries be comprised of at least six people chosen from venires representing the entire community. *See, e.g.*, Ballew v. Georgia, 435 U.S. 223, 231–32 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 fewer than six persons violate the Constitution); Diamond, *supra* note 255, at 153; *cf.* Dallas, *supra* note 321, at 1402 (arguing that heterogeneous corporate boards are less susceptible to confirmation bias because they harbor more competing views).

counteract the biases and fallacies we have discussed? If not, then, are there other measures that could succeed? The next two Subparts 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 in this Article.[326] 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 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.[327] 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, if making that case requires aggregation of the weight of many bits of "small" evidence, that party will get no uniqueness bounce

326. *See, e.g.*, Lon L. Fuller, *The Adversary System*, *in* TALKS ON AMERICAN LAW 34, 39–40 (H. Berman ed., 2d 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." (internal quotation marks omitted)); John Thibaut et al., Comment, *Adversary Presentation and Bias in Legal Decisionmaking*, 86 HARV. L. REV. 386, 396–97 (1972) (arguing that adversarial procedures counteract decision-maker bias more effectively than inquisitorial procedures).

327. *See* KAHNEMAN, *supra* note 207, at 417 ("[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 206 and accompanying text.

and instead will be disadvantaged by multiple irrelevance fallacies.[328] Indeed, as the next Subparts 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 advantages to recur, systematically favoring the prosecution over the defense.

2. The Biasing Effect of the State's Monopoly over the Initial Investigation

The *DeLuna* case discussed above[329] 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 begging someone to take the store's money, then screaming and struggling. When police arrived moments later, they found her drenched in blood and near death after an assailant had stabbed her and tried to wrestle her into a back room.[330]

A police audiotape captured the 911 call and subsequent radio traffic while police hunted for a man who had been seen fleeing the store on foot.[331] Prosecutors played the 911 tape at DeLuna's trial but falsely informed defense counsel that the manhunt portion of the tape had been recorded over without being saved.[332] When the manhunt tape came to light years after DeLuna was executed, it revealed that police had chased another man along a different path for twenty-five minutes before a call about DeLuna diverted attention to him.[333] 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: 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.[334] 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.[335] The prosecutor branded DeLuna a liar and Hernandez a "phantom,"[336] and the jury heard nothing more about Hernandez or anything about his history of armed convenience store robberies and knife violence, physical resemblance to DeLuna, or his repeated admissions to stabbing the convenience store

328. *See supra* notes 256–320 and accompanying text.
329. *See supra* notes 17–34, 44 and accompanying text.
330. *See* Liebman et al., *supra* note 17, at 910.
331. *Id.* at 736.
332. *Id.* at 949–50.
333. *Id.* at 951–52.
334. *See supra* notes 19–20 and accompanying text.
335. *See supra* note 23 and accompanying text.
336. *See supra* note 25 and accompanying text.

clerk.[337] Nor did the jury learn that at least ten features of the eyewitness's description of the attacker that did not match DeLuna did match Hernandez.[338]

When police arrested DeLuna forty minutes after the crime, they decided against a station-house lineup and took him to the crime scene for a nighttime show-up identification in a poorly lit parking lot. On the ride over, DeLuna told the arresting officer he did not commit the crime but knew who did. Police never followed up or questioned DeLuna.[339] When the squad car arrived at the convenience store, the eyewitness was initially too scared to view the suspect, but relented when officers surrounded him and let him view the seated, shirtless and hand-cuffed suspect through the squad car window while flashlights were trained on his face. After observing DeLuna for fifteen seconds, the Anglo witness identified him as the culprit.[340] Years later, the witness admitted he had trouble telling one Latino from another and said he was only "seventy percent sure" of the identification and would have been only "fifty-fifty" if police had not told him beforehand that they had found DeLuna cowering under a pickup truck.[341]

Meanwhile, a lone detective and a police photographer had entered the store where the killing took place. Evidence the two overlooked, never tested forensically, or failed to disclose to the defense is listed above.[342] The detective found the lock-blade buck knife used to kill the victim, 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.[343] No prints were found, and the graphite spoiled the knife for later lab testing performed with proper materials.[344] None of the prints found elsewhere at the scene matched DeLuna.[345] 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 the employees.[346] They scrubbed it down overnight—wiping away a bloody handprint, unnoticed by police, on the inside of a window—and opened on

337. *See* Mills & Possley, *A Phantom*, *supra* note 17.
338. *See supra* Figure 2.
339. *See* Liebman et al., *supra* note 17, at 766.
340. *See id.* at 762.
341. *See id.* at 765.
342. *See supra* Figure 2.
343. *See* Liebman et al., *supra* note 17, at 925.
344. *Id.*
345. Grieg, *supra* note 17, at 95.
346. *See* Liebman et al., *supra* note 17, at 926.

time the next morning.[347] No one ever saw the intact crime scene in daylight.[348]

These facts suggest how heuristic biases may affect police officers, forensic analysts, and prosecutors when they exercise their monopoly over the initial handling of vast amounts of potential evidence, and the heuristic advantage that 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.[349] Because a case premised on seemingly unique, "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, DNA hit, or fingerprint match. The *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 lineup 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 and spotlighted his face with flashlights.[350]

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, particularly because such evidence might *not* corroborate the "big" evidence and, therefore, might weaken the state's case. Investigation that does take place tends, at least subconsciously, to "confirm" what the "big" evidence has simulated and "represents," namely the suspect's guilt.[351] 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.[352]

In *DeLuna*, for example, when the detective ended her scene investigation upon hearing of the "big" show-up identification, she left

347. *See id.* at 912.
348. *See id.* at 908–27.
349. *See supra* note 208 and accompanying text.
350. *See* Liebman et al., *supra* note 17, at 761–66.
351. *See supra* notes 230–36 and accompanying text.
352. *See supra* notes 287–305 and accompanying text.

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 whom no one interviewed at the time and several items spotted in previously undisclosed police photographs:

- 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. No mention was made at trial of the absence of foreign hair in DeLuna's fingernails and on his clothing at the time of his arrest.[353]
- Bloody shoeprints. The victim was barefoot, so bloody shoeprints found inside the store and on the sidewalk outside the store 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.[354]
- 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.[355]
- 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.[356]

Also unexplored were several non-exclusionary non-matches that had appeared before the scene investigation ended: DeLuna's claim that he had seen another man commit 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.[357]

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

353. *See* Liebman et al., *supra* note 17, at 914, 921 fig.29-B.

354. *See id.* at 917–21 & figs.27, 29-A, 29-B.

355. *See id.* at 916–17 & fig.25.

356. *See id.* at 916 fig.25; *supra* Figure 2.

357. *See* Liebman et al., *supra* note 17, at 731–49, 775–82, 908–38; *supra* Figure 2.

professionals to assign to the scene investigation, how long and thoroughly to search, how best to lift fingerprints, and what evidence to preserve and test. Eyewitnesses also will have been solely at the state's disposal 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 shoe prints and a chance to apprehend and question alternative suspects. Unique access to the crime scene and unilateral control over the decision to end the investigation there thus will have maximized the state's access to "big" evidence and heuristically powerful scenarios, and let it choose when its theory likely "covered"[358] the highest possible proportion of found "small" evidence, deterring a search for more. In important respects, the state's monopoly over the initial investigation will have given it the exclusive power to determine whether much of the evidence even *exists*—including "big" evidence pointing *away* from the state's prime suspect and a plethora of "small" evidence.

Conversely, the defense may never have a shot at the crime scene or witnesses that police failed to identify, and the defense will speak to many found witnesses only after police have shaped their belief about what happened.[359] The only scenario-generating evidence to which an innocent defendant is assured access is whatever heuristically disadvantageous "small" non-matches law enforcement happened to discover before shutting down the investigation and decided to disclose under discovery rules built to apply only to "big" evidence.[360] 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.[361]

There are other heuristic skews as well. The very job of a detective and prosecutor is to simulate a "bad guy's" guilt and to rely, 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 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 a defense counsel need embrace is

358. *See* Heller, *supra* note 36, at 296; *supra* note 222 and accompanying text.

359. *See, e.g., supra* notes 340–41 and accompanying text. State's witnesses are not obliged to speak to defense counsel and are often advised of that fact by police officers who themselves typically decline to talk to the defense.

360. *See infra* Part V.C.3.

361. *See supra* note 228 and accompanying text.

reasonable doubt. Because defense lawyers trade in the pallid probabilities necessary to avoid conviction, there is less chance that 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 is, then, heuristic truth behind the 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, rather than a comparison of all possible suspects, they present decision makers with a single, prominent theory encapsulated in a prior arrest warrant or indictment and in the state's opening arguments, case-in-chief, and summation. 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 that the adversarial system presents has only two possible answers: either the defendant did it or the culprit escaped justice. Of those options, 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 decision maker 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, therefore, 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 hardwiring 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, it can deter jurors from taking shortcuts and focus them intuitively on counterweights to biases, such as alternative scenarios, base rates, and denominator values slightly higher than numerators. As we have seen, however, much of the heuristic damage occurs during the initial investigation before the adversarial system kicks in. Worse yet, even true "experts," and surely lay jurors, succumb to the heuristic homunculus that jumps up and down even in Stephen Jay Gould's head.[362]

362. *See supra* note 266 and accompanying text; *see also supra* note 211 and accompanying text (noting that even scientists and other "experts" are susceptible to heuristic biases).

The question remains whether other existing legal doctrines, such as the evidentiary rules mentioned above,[363] achieve a proper balance, knocking "big," seemingly "unique" evidence down to size and adding clout to seemingly "irrelevant" non-exclusionary non-matches. In the next Subpart, 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. Far from curing the heuristic inflation of "big" evidence and deflation of aggregations of "small" evidence, however, 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 nearly all jurisdictions draw a related distinction between "direct" and "circumstantial" evidence.[364] For example, 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 defined as "indirect evidence . . . from which you could find that another fact exists, even though it has not been proved directly."[365] 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."[366]

Following the modern trend, both Connecticut's and California's instructions (in the words of Connecticut's) say there is "no legal distinction between direct and circumstantial evidence as far as probative value" and let

363. *See supra* notes 270–79 and accompanying text.

364. *See* 1 BARBARA E. BERGMAN & NANCY HOLLANDER, WHARTON'S CRIMINAL EVIDENCE § 1:8 (15th ed. 1997) (collecting instructions distinguishing "direct" and "circumstantial" evidence).

365. Connecticut Criminal Jury Instructions § 2.4-1 (4th ed. 2010), *available at* http://www.jud.ct.gov/JI/criminal/.

366. Judicial Council of California Criminal Jury Instructions (CALCRIM) No. 223 (2012) [hereinafter CALCRIM], *available at* http://www.courts.ca.gov/partners/documents/calcrim_juryins.pdf.

jurors "give equal weight to both."[367] But both still tend to confuse jurors and convey the opposite message. Both juxtapose the encomium "direct" with 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*")[368] and more powerful ("prov[ing] a fact by itself" versus "not directly prov[ing] the fact"[369]).[370] No wonder, then, that jurors continue to associate "direct" with "strong" and "circumstantial" with "weak" evidence.[371] In one study, subjects presented with scenarios involving strong circumstantial evidence or weak direct evidence and then given California's direct-versus-circumstantial instruction categorized the evidence incorrectly 45% to 85% of the time.[372] Subjects incorrectly defined evidence as direct or circumstantial 38% of the time when given modern instructions like Connecticut's and California's, and 49% of the time when given old-fashioned, jargon-filled instructions.[373] Worse, a substantial minority of American jurisdictions invites jurors to privilege "direct" over "circumstantial" evidence by defining direct evidence as the norm and circumstantial evidence as something the law merely tolerates, giving jurors lengthy directions about how to decide whether to accept circumstantial evidence as proof, or failing to instruct jurors that circumstantial evidence can be as weighty as direct evidence.[374]

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.[375] 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

367. Connecticut Criminal Jury Instructions, *supra* note 365, § 2.4-1; *see also* CALCRIM, *supra* note 366, No. 223 ("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 225, at 1803 nn.6–7 (citing federal and state case law affirming this proposition).

368. Connecticut Criminal Jury Instructions, *supra* note 365, § 2.4-1 (emphasis added).

369. CALCRIM, *supra* note 366, No. 223.

370. *See* Tiersma & Curtis, *supra* note 103, at 257 (presenting study results indicating that all direct versus circumstantial instructions risk reinforcing the "popular misconception that circumstantial evidence is weak").

371. *Id.* at 246–56 (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 some jurors to treat direct evidence as stronger and more reliable, and circumstantial evidence as weaker).

372. *Id.* at 251–56, 259–61.

373. *See id.* at 250.

374. *See* Greenstein, *supra* note 225, at 1803–04 & n.7 (citing examples and cases).

375. *See supra* notes 46, 95–123 and accompanying text.

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 decision makers into the uniqueness fallacy in favor of "big" evidence and the irrelevance fallacy against "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 before all potential clues and witnesses have been queried, with results preserved 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 common scenarios—those 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, and the latter "weak" evidence further reinforces the uniqueness fallacy. What then does the law do to extend the period during which evidence is collected, 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, and bring the case to the district attorney to prosecute—are either entirely within the discretion of law enforcement[376] or are subject only to a weak requirement of probable cause.[377] Few of the city and county departments that make these decisions have guidelines on these matters, and the guidance that does exist is rarely public or enforceable. State courts and attorneys general typically honor

376. *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." (quoting Bordenkircher v. Hayes, 434 U.S. 357, 364 (1978)) (internal quotation marks omitted)); United States v. Cox, 342 F.2d 167, 172 (5th Cir. 1965) (similar).

377. *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.").

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.[378]

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."[379] Except for "chain of custody" rules, which apply only when the state confiscates and introduces evidence at trial,[380] 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 *Arizona v. Youngblood*, 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.[381] By withholding redress absent outright bad faith, *Youngblood* "imposes an almost insurmountable burden upon [an] accused" harmed by law enforcement misfeasance[382] 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.[383] Federal and most state courts continue to apply *Youngblood*.[384] Even the handful of

378. *See, e.g.*, Connick v. Thompson, 131 S. Ct. 1350, 1355–56, 1366 (2011) (granting a prosecutor immunity from damages after the prosecutor failed to inform a falsely convicted defendant of a pretrial forensic test showing that the killer's blood type excluded the defendant who had spent years on death row); Van de Kamp v. Goldstein, 129 S. Ct. 855, 858–59, 863–64 (2009) (granting absolute immunity to a district attorney responsible for breakdowns in supervision, training, and information management that led to a failure to disclose exculpatory information against a plaintiff who spent twenty-four years in prison for a murder he did not commit); Joseph v. Yocum, 53 F. App'x 1, 3 (10th Cir. 2002) ("[Prosecutors] are absolutely immune from liability . . . for the 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, 1411 (3d Cir. 1991) ("A prosecutor's alleged failure to properly investigate before initiating a prosecution is . . . within the scope of absolute immunity.").

379. 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, No. H-07-2087, 2008 WL 4425366, at *7 (S.D. Tex. Sept. 24, 2008)) (internal quotation marks omitted).

380. *See* LEMPERT ET AL., *supra* note 99, at 1216–21 (discussing chain of custody).

381. Arizona v. Youngblood, 488 U.S. 51, 58 (1988) (holding that state destruction of evidence potentially favorable to the defense did not violate due process in the absence of bad faith). Other cases extend *Youngblood* to bar relief for law enforcement's failure to collect evidence. *See, e.g.*, State v. Krosch, 642 N.W.2d 713, 719 (Minn. 2002); State v. Ware, 881 P.2d 679, 683–84 (N.M. 1994).

382. Norman C. Bay, *Old Blood, Bad Blood, and* Youngblood*: Due Process, Lost Evidence, and the Limits of Bad Faith*, 86 WASH. U. L. REV. 241, 247 (2008).

383. *See id.* at 243–44; Peter Neufeld, *Legal and Ethical Implications of Post-Conviction DNA Exonerations*, 35 NEW ENG. L. REV. 639, 646 (2001).

384. *See, e.g.*, Cost v. State, 10 A.3d 184, 193 (Md. 2010) (citing cases); People v. al-Azawi, No. 292425, 2010 WL 4026107, at *4 (Mich. Ct. App. Oct. 14, 2010) (per curiam) ("A

states that have modified it only give judges discretion to instruct jurors that they may draw adverse inferences against the state if officials destroyed or failed to secure important forensic evidence for no good reason and if the state's case is otherwise weak.[385]

Neither the accused nor jurors even have a right to know about defects in the state's investigation. Courts rarely find *Brady* or other discovery violations when police or prosecutors fail to reveal defects,[386] and absent bad faith, they do not treat a failure to discover or collect evidence as a *Brady* violation in its own right.[387] 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 to introducing evidence the defense has found that the state could have, but did not, obtain.[388] If the defense manages to expose weaknesses in the

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 382, at 247 n.17, 287 n.364 (citing cases).

385. *See, e.g.*, *Cost*, 10 A.3d at 188 n.3, 192–93, 196–97 (citing cases and adopting a rule allowing the jury to draw adverse inferences if the state "destroyed highly relevant evidence in its custody that it normally would have retained and submitted to forensic examination"); Commonwealth v. Kee, 870 N.E.2d 57, 65 (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 382, at 287–89 & nn.364–75 (citing cases from states allowing adverse-inference instructions). Several states have also recently adopted laws requiring preservation of DNA evidence. *See id.* at 246 & nn.15–16, 284–85 & nn.351–58.

386. *See, e.g.*, Woodruff v. State, 608 N.W.2d 881, 888 (Minn. 2000) (finding no *Brady* violation despite the state's failure to disclose evidence supporting the 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 7, 2010) (holding that the failure to disclose a police report did not violate *Brady*, though it may have supported the defendant's inadequate-investigation claims). *But see* Workman v. Commonwealth, 636 S.E.2d 368, 376–78 (Va. 2006) (finding that the failure to disclose evidence supporting an inadequate-investigation claim violated *Brady*).

387. *See, e.g.*, State v. Ware, 881 P.2d 679, 684–85 (N.M. 1994) (surveying cases and holding that only evidence missed in bad faith can violate the *Brady* rule).

388. *See, e.g.*, Harrington v. Richter, 131 S. Ct. 770, 782 (2011) (noting that the defense at trial "stressed deficiencies in the [police and forensic] investigation"); Crawford v. Head, 311 F.3d 1288, 1302–05 (11th Cir. 2002) (noting inadequate-investigation claims); Moreno v. State, No. A-10041, 2010 WL 5021778, at *3 (Alaska Ct. App. Dec. 8, 2010) (finding that the defense had a fair opportunity to criticize the police investigation); State v. Martinez, No. 2 CA-CR 2007-0355, 2010 WL 3928631, at *8 (Ariz. Ct. App. Sept. 30, 2010) (noting inadequate-investigation claims); People v. Corbett, Nos. H032848, H035224, 2011 WL 18733, at *13–14 (Cal. Ct. App. Jan. 5, 2011) (same); State v. Jamison, 742 A.2d 1272, 1274 (Conn. App. Ct. 1999) (same); State v. Cosey, 1997-2020, p. 13 (La. 11/28/00); 779 So. 2d 675, 684 (same); Fullbright v. State, 895 A.2d 1088, 1095–96 (Md. Ct. Spec. App. 2006) (same); People v. Hill, No. 287226, 2010 WL 539841, at *1 (Mich. Ct. App. Feb. 16, 2010) (per curiam) (same); King v. State, 2008-KA-01509-SCT (¶¶16–17, ¶20) (Miss. 2010); 47 So. 3d 658, 665 (same); State v. Campobasso, No. 303042-II, 2004 WL 2284208, at *1 (Wash. Ct. App. Oct. 12, 2004) (same); *cf.* Blankenship v. Hall, 542 F.3d 1253, 1270–74, 1281 (11th Cir. 2008) (rejecting an

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.[389] With the sole exception of Massachusetts,[390] no jurisdiction recognizes an inadequate-investigation defense or requires judges to instruct jurors that they may treat inadequacies in the state's investigation as sufficient in themselves to establish reasonable doubt.

Even in Massachusetts, officers have no duty to gather exculpatory evidence.[391] Trial judges are simply allowed, but not required,[392] 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."[393] 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.[394]

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 will not even know about problems plaguing police investigations,

---

ineffective-assistance-of-counsel claim based on the failure to present a convincing insufficient-investigation defense), *cert. denied*, 131 S. Ct. 1041 (2011); State v. Herring, 762 N.E.2d 940, 957–58 (Ohio 2002) (similar).

389. *See, e.g.*, Morris v. State, 2000-CT-0822-SCT (¶13) (Miss. 2006); 927 So. 2d 744, 748 ("[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."); Lane v. State, Nos. 14-07-00871-CR, 14-07-00872-CR, 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.").

390. *See, e.g.*, Commonwealth v. Bowden, 399 N.E.2d 482, 491 (Mass. 1980) (recognizing a defense based on the "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"). For Massachusetts cases applying the *Bowden* defense, see, for example, Commonwealth v. Silva-Santiago, 906 N.E.2d 299, 315 (Mass. 2009) ("[I]nformation regarding a third-party culprit, whose existence was known to the police but whose potential involvement was never investigated, may be admissible under a *Bowden* 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 . . . and . . . such failures alone may be sufficient to create a reasonable doubt of the defendant's guilt."); *cf.* Massachusetts Criminal Model Jury Instructions No. 3.740 (2009), *available at* http://www.mass.gov/courts/courtsandjudges/courts/districtcourt/jury-instructions/criminal/pdf/3740-evidence-omissions-in-police-investigations.pdf.

391. *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.").

392. *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).

393. Massachusetts Criminal Model Jury Instructions, *supra* note 390, No. 3.740.

394. *Id.*

and only 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 seemingly unique evidence of identity and overcome its heuristically induced blindness to individually "small," though conjointly powerful, bases for reasonable doubt.

### 3. Rules Regulating Discovery

Even when the state obtains evidence of non-exclusionary non-matches, discovery rules invite the state to withhold the information from the defense on the ground that non-matches are individually irrelevant. This is particularly true after officials have committed themselves to an inculpatory scenario built around "big" evidence. In essence, the *Brady* rule defining the constitutional duty to disclose builds the representativeness, simulation, and confirmation biases into its own operation.[395]

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 the view that prosecutions are a poker game in which holding one's losing cards close is an acceptable strategy for success.[396] The Court recognized a justice interest in state disclosure[397] of evidence that is both "exculpatory" because it makes it more likely than without the evidence that the defendant is innocent, and "material" in that nondisclosure had some effect on the outcome.[398] The

395. *See* 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 . . . ."). In *United States v. Bagley*, the Court held that undisclosed evidence is material "if there is a reasonable probability that, had the evidence been disclosed . . . the result of the proceeding would have been different." United States v. Bagley, 473 U.S. 667, 682 (1985).

396. *See Brady*, 373 U.S. at 86–88; *see also* 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.

397. *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.").

398. *See Brady*, 373 U.S. at 87–88 ("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 helps shape a trial that bears

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.[399] 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.[400]

A problem with the *Brady* rule is that it does not tell prosecutors what they are expected to disclose *before trial*. It includes no requirement that prosecutors "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. As Professor Alafair Burke notes, *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."[401]

Bizarrely, *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 about the court's *future* ruling.[402] Worse yet, by far the most likely outcome 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 all five of the following 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

heavily on the defendant. That casts the prosecutor in the role of an architect of a proceeding that does not comport with . . . justice . . . .").

399. *See id.* at 84–85, 91.

400. *Bagley*, 473 U.S. at 682.

401. *See* Alafair Burke, Commentary, Brady*'s Brainteaser: The Accidental Prosecutor and Cognitive Bias*, 57 CASE W. RES. L. REV. 575, 576 (2007).

402. *See, e.g.*, Burke, *Improving Prosecutorial Decision Making, supra* note 230, at 1610–12 (criticizing *Brady* because "applying the 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 v. Maryland*: Games Prosecutors Play*, 57 CASE W. RES. L. REV. 531, 558–59 (2007) (criticizing the post-hoc evaluation that the *Brady* rule requires).

appeals on that ground; and (5) the reviewing court finds that the withheld evidence was "material."

Condition (1) is rare, to begin with, because over 90% of prosecutions end in plea bargains.[403] Indeed, the prosecutor at least marginally increases the chances of a guilty plea by withholding exculpatory evidence that might embolden the defendant to go to trial. Condition (3) is also rare because the prosecution controls the evidence it withheld, making it hard for defendants to discover. Finally, reversals of criminal verdicts on any grounds, much less on *Brady* grounds—condition (5)—occur in only a small percentage of the miniscule subset of cases meeting the other conditions.[404] *Any* appellate determination, especially a reversal, is unlikely, which makes it sensible for conscientious prosecutors to err on the side of withholding evidence. As Justice Marshall noted, the "materiality" standard gives prosecutors the

> impossible task of deciding whether a certain piece of information will have a significant impact on the trial, bearing in mind that a defendant will later shoulder the heavy burden of proving how it would have affected the outcome. At best, this standard places on the prosecutor a responsibility to speculate, at times without foundation, since the prosecutor will not normally know what strategy the defense will pursue or what evidence the defense will find useful. At worst, the standard invites a prosecutor, whose interests are conflicting, to gamble, to play the odds, and to take a chance that evidence will later turn out not to have been potentially dispositive.[405]

Still worse, when prosecutors predict, speculate, and play the odds, heuristic biases assure that they will identify few pieces of exculpatory evidence—particularly non-exclusionary non-matches—that have to be disclosed.[406] 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

403. *See* RONALD JAY ALLEN ET AL., COMPREHENSIVE CRIMINAL PROCEDURE 1161 (2d ed. 2005) ("Of felony convictions nationwide, 94 percent are obtained by guilty plea.").

404. *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 (1990–1991) (finding 3% reversal rate in 1970s federal habeas corpus cases in the Southern District of New York).

405. *Bagley*, 473 U.S. at 701 (Marshall, J., dissenting).

406. *See* Burke, *supra* note 401, at 576–77 (arguing that cognitive biases triggered by *Brady* "invite[] prosecutors to systematically undervalue the materiality of evidence").

the scenario, say, through eyewitness testimony identifying the defendant or through the defendant's own admissions. The jury finds the state's theory is 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.[407]

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

- 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.[408]
- 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.[409]
- The confirmation bias. In imagining an appellate decision as to whether the withheld evidence would have changed the outcome, she will bring to mind the strong parts of her case and the weaknesses in the evidence under consideration.[410]
- An outcome bias twice over.[411] 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 to assume innocent explanations for non-exclusionary non-matches, the above biases will create a false

407. The *Brady* rule requires reviewing courts to make an "independent examination of the record," *Bagley*, 473 U.S. at 679 n.8 (Blackmun, J., joined by O'Connor, J.), 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, 556 U.S. 449, 469–70 (2009); Youngblood v. West Virginia, 547 U.S. 867, 870 (2006).

408. *See supra* Part V.A.3.a.

409. *See supra* Part V.A.2.

410. *See* Burke, *supra* note 401, 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* Part V.A.3.b.

411. Outcome bias leads decision makers to treat the very fact of an outcome as retroactively validating poor decisions producing the outcome. *See, e.g.*, KAHNEMAN, *supra* note 207, at 203–04; Jonathan Baron & John C. Hershey, *Outcome Bias in Decision Evaluation*, 54 J. PERSONALITY & SOC. PSYCHOL. 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").

> certainty in the uniqueness of the former and irrelevance of the latter.[412]

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 already 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 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's 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."[413] If Smith could present evidence establishing a counter-scenario, with a different suspect in the role of villain, heuristic problems could be allayed.[414] But the realities of our criminal justice system and the 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.

In reality, few defendants have meaningful access to evidence implicating third-party suspects, especially innocent defendants who are disconnected from the events charged. 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,[415] the absence of a duty to conduct police investigations competently or face consequences,[416] heuristic and legal limits on officials' obligation to turn over leads that do not support the state's case,[417] chronic underfunding of indigent criminal-defense investigations,[418] defense counsel's disposition to

412. *See supra* Part V.A.4–5.

413. *See supra* notes 58–59 and accompanying text; text accompanying notes 361–62.

414. *See* Heller, *supra* note 36, at 290–98 (showing that "priming" with alternative, exculpatory scenarios can overcome unfair heuristic advantages accompanying the state's "direct" evidence).

415. *See supra* notes 328–61.

416. *See supra* Part V.C.2.

417. *See supra* Part V.C.3.

418. *See, e.g.*, STANDING COMM. ON LEGAL AID & INDIGENT DEFENDANTS, AM. BAR ASS'N, *GIDEON*'S BROKEN PROMISE: AMERICA'S CONTINUING QUEST FOR EQUAL JUSTICE 10–11, 38

give priority in deploying limited resources to poking holes in the state's case and not to constructing her own, and the disposition of most criminal cases by plea bargains aimed at minimizing the cost of investigations.[419]

Further decreasing incentives to look for evidence of a third-party's guilt are the 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 once the defendant presents any evidence suggesting that someone else committed the crime—an alibi, for example—the state must dispel the implication of that evidence beyond a reasonable doubt.[420] Typically, the defendant can trigger this process with any evidence that is relevant and not substantially more prejudicial than probative.[421] 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.[422] The defendant's Sixth Amendment right to present defensive evidence undergirds these principles.[423]

But this is not how the law works when it comes to the defense that it was Jones, not the defendant Smith, who committed the crime. As the Supreme Court noted 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.[424] In these jurisdictions, such evidence is inadmissible unless it

(2004), *available at* http://www.americanbar.org/content/dam/aba/administrative/legal_aid_indigent_defendants/ls_sclaid_def_bp_right_to_counsel_in_criminal_proceedings.authcheckdam.pdf.

419. *See* ALLEN ET AL., *supra* note 403, 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).

420. *See, e.g.*, MODEL PENAL CODE §§ 1.12(1)–(2)(a), 1.13(9)(c) (1985) (providing that once evidence negating an element of the offense is offered, the state must disprove that defense beyond a reasonable doubt).

421. *See, e.g.*, FED. R. EVID. 401–403.

422. *See id.* 401 advisory committee's note ("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' . . . ."); Bourjaily v. United States, 483 U.S. 171, 179–80 (1987) ("[I]ndividual pieces of evidence, insufficient in themselves to prove a point, may in cumulation prove it.").

423. *See, e.g.*, Crane v. Kentucky, 476 U.S. 683, 690–91 (1986); Chambers v. Mississippi, 410 U.S. 284, 294–97 (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 . . . .").

424. Holmes v. South Carolina, 547 U.S. 319, 327 & n.* (2006) (citing cases adopting "rules regulating the admission of evidence proffered by criminal defendants to show that someone else committed the crime"); *see also* 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." (quoting Smithart v.

(1) establishes a "clear link" or "direct connection" between the alternate suspect and the crime (the majority rule);[425] (2) satisfies the defense's effective standard of *proof* by singlehandedly establishing a reasonable doubt;[426] or (3) is more probative than prejudicial, reversing the usual presumption in favor of admissibility.[427] Applying these doctrines, courts have excluded evidence that police initially arrested another man for the crime;[428] that the father of a child whom the defendant was accused of sexually abusing had himself been convicted of sexual abuse some time thereafter;[429] that a man, whom the trial court forbade the defendant to call at trial, was identified by two paramedics as resembling a male they saw at the crime scene acting suspiciously and then fleeing;[430] that police obtained statements from a man with a motive to commit the crime who revealed an unexplained familiarity with the details of the crime;[431] and that a man with no connection to the defendant had been found in possession of a distinctive box resembling one taken from the victim around the time of the crime.[432] Only when evidence implicating a third-party suspect verges on

State, 988 P.2d 583, 586–87 (Alaska 1999)) (internal quotation marks omitted)); 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 (1996) (criticizing majority rule placing a "direct connection" limitation on evidence of a third party's guilt; also criticizing alternative rules that (1) impose a "capable-of-raising-a-reasonable-doubt" standard, which is stricter than the usual balance of probative weight and prejudice; or (2) apply the usual "probative weight versus prejudice" balance in an unusually strict manner when evaluating evidence of a third party's guilt).

425. *See, e.g.*, McCord, *supra* note 424, at 936–38 & n.99 (citing cases).

426. *See, e.g.*, 41 C.J.S. *Homicide* § 318 (2012) ("Evidence tending to show the commission by another person of the crime charged may be introduced by accused [only] when it is inconsistent with, and raises a reasonable doubt of, his own guilt . . . ."); McCord, *supra* note 424, at 936–38 & n.100 (citing cases).

427. *See* McCord, *supra* note 424, at 936–38 & n.101 (citing cases).

428. State v. Oliver, 821 P.2d 250, 251–52 (Ariz. Ct. App. 1991) (rejecting alternative-suspect evidence absent an "inherent tendency to connect the other 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).

429. People v. Sparman, 608 N.Y.S.2d 672, 673–74 (App. Div. 1994) (excluding evidence implicating an alternate suspect with greater access to the sexual-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" (quoting People v. Brown, 590 N.Y.S.2d 896, 898 (App. Div. 1992)) (internal quotation marks omitted), *granting habeas sub nom.* Sparman v. Edwards, 26 F. Supp. 2d 450 (E.D.N.Y. 1997), *aff'd*, 154 F.3d 51 (2d Cir. 1998)).

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

431. State v. Gilman, 608 A.2d 660, 663 (Vt. 1992) (barring alternative-suspect evidence unless "motive and opportunity have been shown" and it incriminates the third party "directly" (quoting State v. Denny, 357 N.W.2d 12, 17 (Wis. Ct. App. 1984))).

432. State v. McNeill, 392 S.E.2d 78, 83–84 (N.C. 1990) (barring alternate-suspect evidence that does not "point directly to the guilt of some specific person" and is "not inconsistent with the defendant's own guilt"). For a discussion of other courts' treatment of alternate-suspect evidence, see McCord, *supra* note 424, at 950–51 & n.167.

"big" evidence[433]—DNA that excludes the defendant and is consistent with the alternate suspect or an eyewitness identification of that suspect—is the court likely to let the defendant introduce the evidence.[434]

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 . . . strong forensic evidence, the proffered evidence about a third party's alleged guilt' may (or perhaps must) be excluded."[435] Despite this ruling, the Court seemed to approve the rule that most other states apply: "[E]vidence 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 . . . ."[436] 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 that match an alternative suspect.[437] 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. Thus, while modestly mitigating the uniqueness fallacy, *Holmes* bolsters the irrelevance fallacy.

#### 5. Rules Limiting Statistical Evidence

In *Collins*, 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

433. *See, e.g.*, *McNeill*, 392 S.E.2d at 83 ("[E]vidence must (1) point directly to the guilt of some specific person, and (2) be inconsistent with the defendant's guilt.").

434. *See, e.g.*, United States v. Armstrong, 621 F.2d 951, 953 (9th Cir. 1980) (finding that the trial court improperly excluded the defendant's evidence that an alternate suspect had bait money from the robbery with which the defendant was charged); United States v. Robinson, 544 F.2d 110, 112–13 (2d Cir. 1976) (reversing a conviction because the trial court excluded evidence that a correctional officer identified someone other than the defendant as a bank robber in a surveillance videotape recording); Joyner v. State, 678 N.E.2d 386, 389–90 (Ind. 1997) (affirming the admission of mitochondrial DNA evidence of a hair found on the victim that excluded the defendant and likely matched the alternate suspect); McCord, *supra* note 424, at 951 & n.168 (citing similar cases).

435. Holmes v. South Carolina, 547 U.S. 319, 329 (2006) (alteration in original) (emphasis added) (quoting State v. Holmes, 605 S.E.2d 19, 24 (S.C. 2004)).

436. *Id.* at 328 (alteration in original) (quoting State v. Gregory, 16 S.E.2d 532, 534 (S.C. 1942)) (internal quotation marks omitted).

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

who, like the defendant couple, matched the perpetrators' description.[438] As we note above, however, rather than limit its reversal to the expert's manifold technical mistakes,[439] the court suggested 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."[440] 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 distinctive characteristics could be found in the entire Los Angeles area."[441] What we now know about heuristic bias—particularly the certainty effect—undermines both assumptions. Contrary to the first assumption, human decision makers' quest for certainty leads them systematically to undervalue explicit probabilities, not to overvalue them.[442] 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—for example, the eyewitness identification of Carlos DeLuna, 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.[443]

Despite these flaws, many courts have relied on *Collins*'s reasoning as elucidated by Professor Tribe as a basis for barring aggregative statistical analysis.[444] Citing *Collins*, the Minnesota Supreme Court has gone so far as to

438. People v. Collins, 438 P.2d 33, 36–42 (Cal. 1968).

439. *See id.* at 39; *supra* notes 79–80 and accompanying text.

440. *Collins*, 438 P.2d at 40.

441. *Id.*

442. *See supra* Part V.A.3.c.

443. *See supra* notes 17–44 & Figure 2, 314–20, 329–57 and accompanying text.

444. *See, e.g.*, United States v. Gwaltney, 790 F.2d 1378, 1383 (9th Cir. 1986) (citing *Collins* as the standard for admitting probabilistic evidence); United States *ex rel.* DiGiacomo v. Franzen, 680 F.2d 515, 518 (7th Cir. 1982) (citing *Collins* to support the view that probabilities confuse jurors); United States v. Yee, 134 F.R.D. 161, 211 (N.D. Ohio 1991) (citing *Collins* in support of the view that probabilistic testimony is often ruled inadmissible), *aff'd sub nom.* United States v. Bonds, 12 F.3d 540 (6th Cir. 1993); Hart v. Sec'y of the Dep't of Health & Human Servs., 60 Fed. Cl. 598, 606 (2004) (citing *Collins* in invalidating a special master's reliance on probabilistic analysis to rule out a vaccine as the cause of an infant's death); 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, 134–35 (Ind. Ct. App. 1985) (citing *Collins* in barring probabilistic evidence absent "empirical scientific data"); Wilson v. State, 803 A.2d 1034, 1044–46 (Md. 2002) (barring use of the multiplication rule in

impose a blanket ban on even admittedly sound probabilistic estimates of the frequency of identifying traits in the population that are offered as an aid in adjudicating identity in criminal cases.[445] Even when the court created an exception for DNA evidence,[446] acknowledging 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, it still declined to relax its general ban on statistical quantification and aggregation.[447] Although most courts have not gone as far as the Minnesota Supreme Court, nearly all of them 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 regarding the number of possible suspects (relevant to calculating the prior odds of guilt); (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.

assessing the likelihood of two infants in the same family dying of sudden infant death syndrome (SIDS)); Commonwealth v. Drayton, 434 N.E.2d 997, 1005 (Mass. 1982) (disallowing a fingerprint expert's estimate of "the 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 the "landmark" *Collins* decision in invalidating statistics as a deficient basis for concluding that intentional asphyxiation, not SIDS, caused a child's death); Pearson v. State, 811 P.2d 704, 707–08 (Wyo. 1991) (citing *Collins* in excluding probabilistic evidence); *Adams I*, [1996] 2 Crim. App. 467 at 482 (Eng.) ("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* Part II.B); Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 J. LEGAL STUD. 107, 136 n.38 (2007) (arguing that Bayesian analysis is of little value in reaching reliable trial outcomes); Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439, 442 (1986) ("Among legal academics it is generally agreed that Tribe won this particular debate."). For more sympathetic views of the value of using Bayesian and other types of statistical analysis in resolving factual disputes in court, see, for example, Lempert, *supra* at 442; Nance, *supra* note 124, at 1595–616; Posner, *supra* note 46, at 1508.

445. *See* State v. Joon Kyu Kim, 398 N.W.2d 544, 547–48 (Minn. 1987) (excluding expert testimony that semen found in the victim's body and on a bed was consistent with Kim's blood type and that 96.4% of males in the Twin Cities could be excluded as possible sources of the semen); State v. Boyd, 331 N.W.2d 480, 481–83 (Minn. 1983) (barring evidence of a 99.9% probability of paternity based on blood tests); State v. Carlson, 267 N.W.2d 170, 176 (Minn. 1978) (concluding that although probabilities offered by a 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.'").

446. *See, e.g.*, State v. Bloom, 516 N.W.2d 159, 167 (Minn. 1994) ("[A] DNA exception to the rule against admission of quantitative, statistical probability evidence in criminal prosecutions to prove identity is justified.").

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

To be sure, gaps have appeared in the *Collins* consensus. 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).[448] Nearly all courts 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 outcomes for different estimates of prior odds, numerator, and denominator (step 3).[449] With limited exceptions, however, courts continue to resist evidence regarding 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. This presents a barrier "so extreme that [it] would eliminate the use of statistical evidence" even for DNA and other forensic evidence.[450] Likewise, uses of the multiplication rule (step 2) and Bayesian analysis (steps 1–3) are uncommon outside DNA cases,[451] and we know of no instances in which these techniques have been used to systematically aggregate the force of multiple non-exclusionary non-matches in order to establish a reasonable doubt. To this extent, *Collins* remains the norm, barring the most direct method of dispelling the uniqueness and irrelevance fallacies.

## VI. SENSIBLE REGULATION OF NON-EXCLUSIONARY NON-MATCHES AS EVIDENCE OF IDENTITY

The forces described here are hardwired 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

448. *See, e.g.*, People v. Mountain, 486 N.E.2d 802, 804–06 (N.Y. 1985) (reversing prior precedent and allowing evidence of blood-type matches and associated frequencies); sources cited *supra* note 124.

449. *See, e.g.*, T.A.T. v. R.E.B. (*In re* Paternity of M.J.B.), 425 N.W.2d 404, 408–09 (Wis. 1988) (describing different statistical techniques); *see also supra* notes 122–23, 179 and accompanying text.

450. Koehler, *supra* note 124, at 392. For a general discussion of this barrier and why it exists, see *id.* at 380–93.

451. *See* Nance, *supra* note 124, at 1612 ("Neither the likelihood ratio format nor the chart format is commonly employed at this time in criminal cases in the United States although they do appear in civil paternity cases . . . .").

innovation can occur 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.[452] Part of the literature, emanating from computer scientists and other experts predominantly from the United Kingdom, focuses on ways to help jurors and other criminal-process actors make effective use of Bayesian analysis,[453] including in assessing the conjoint effect of non-exclusionary non-matches.[454] Two of these experts, Norman Fenton and Martin Neil, take as their starting point that "for many people—and this includes . . . highly intelligent barristers, judges and surgeons—any attempt to use Bayes Theorem . . . is completely hopeless."[455] Judges and lawyers "cannot be expected to follow even the simplest instance of Bayes Theorem in its formulaic representation" and "simply switch-off at the sight of a formula."[456] As Felton and Neil's

452. *See* sources cited *supra* notes 321–24.

453. *See, e.g.*, Colin Aitken et al., *A Graphical Model for the Evaluation of Cross-Transfer Evidence in DNA Profiles*, 63 THEORETICAL POPULATION BIOLOGY 179 (2003); A. Biedermann & F. Taroni, *Bayesian Networks and Probabilistic Reasoning About Scientific Evidence When There Is a Lack of Data*, 157 FORENSIC SCI. INT'L 163 (2006); Leda Cosmides & John Tooby, *Are Humans Good Intuitive Statisticians After All? Rethinking Some Conclusions from the Literature on Judgment Under Uncertainty*, 58 COGNITION 1 (1996); A.P. Dawid, *Bayes's Theorem and Weighing Evidence by Juries*, *in* BAYES'S THEOREM 71 (Richard Swinburne ed., 2002); A. Philip Dawid & Ian W. Evett, *Using a Graphical Method To Assist the Evaluation of Complicated Patterns of Evidence*, 42 J. FORENSIC SCI. 226 (1997); Donnelly, *supra* note 63; Fenton & Neil, *supra* note 133, at 114; Norman Fenton & Martin Neil, *The "Jury Observation Fallacy" and the Use of Bayesian Networks To Present Probabilistic Legal Arguments*, 36 MATHEMATICS TODAY 180 (2000) (revised July 31, 2007) [hereinafter Fenton & Neil, *"Jury Observation Fallacy"*], *available at* http://www.dcs.qmul.ac.uk/researchgp/spotlight/jury_fallacy.pdf, at 2–10; Luke M. Froeb & Bruce H. Kobayashi, *Naive, Biased, Yet Bayesian: 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 LAW PROBABILITY & RISK 275, 281 (2007) ("[Bayesian analysis] can be useful for the purely qualitative analysis of masses of evidence . . . ."); Christian Jowett, *Sittin' in the Dock with the Bayes*, 151 NEW L.J. 201 (2001); Nance & Morris, *supra* note 179, at 429–33 (reporting study findings that although subjects typically undervalue forensic match evidence, the use of a "chart format" showing the effect of different prior odds on the likelihood ratio associated with DNA evidence can substantially reduce error); Nance, *supra* note 124, at 1614–16; Bernard Robertson & Tony Vignaux, *Explaining Evidence Logically*, 148 NEW L.J. 159 (1998).

454. *See* Donnelly, *supra* note 63, at 16 (describing a questionnaire prepared by experts for both sides to help jurors with Bayesian analysis in *Adams I*, 2 Crim. App. 467 (Eng.)); Fenton & Neil, *supra* note 133, at 128 (discussing jurors' use of Bayes' Theorem "when there are multiple pieces of possibly contradictory evidence and interdependencies between them").

455. Fenton & Neil, *supra* note 133, at 128 (emphasis omitted); *see also* Allen & Leiter, *supra* note 88, at 1545 (observing that "lay people make a mess" of probabilistic evidence); Tribe, *supra* note 85, at 1368 (arguing that Bayes' Theorem is "completely opaque to all but the trained mathematician"); sources cited *supra* note 216.

456. Fenton & Neil, *supra* note 133, at 117, 128.

experimental studies have shown, however, "it is the use of abstract probabilities and formulas, rather than the underlying concept, that acts as a barrier to understanding."[457] "When a very simple Bayesian argument is presented visually . . . using [graphic representations of] concrete frequencies people not only generally understand it well, but they can construct their own correct simple calculations."[458]

Consider also a well-known study revealing how susceptible experts—Harvard Medical School faculty and students—are to making the same mistakes lay people do in 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 prevalence of 1 in 1000 and a false positive rate of 5% "*actually* has the disease."[459] Only 18% of the experts gave the correct answer of less than 2% *(1/999 x .95/.05 = .95/49.95 → 1.87%)*; 45% ignored the base rate entirely and answered 95%.[460] When Leda Cosmides and John Tooby replicated the study with Stanford undergraduates, their subjects performed even worse: 12% got the right answer, and 56% ignored the base rate.[461] By taking a few simple steps, however, the researchers made most of the error go away. 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 traits and 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.[462] This reversed the results: 56% got the right answer and only 4% ignored the base rate.[463] By asking a series of "probe questions"—akin to special verdicts—that modeled key steps in correct Bayesian analysis, they increased the number of subjects who got the right answer to 76%, with *none* entirely ignoring the base rate.[464] Even without probe questions, 76% of subjects answered correctly when references to percentages were omitted and frequencies were stated as "how many of 1000 people" fell into particular categories.[465] Finally, when the researchers created a pictorial schema of a grid of tiny boxes representing the "1000 Americans" and asked subjects to answer the

457. *Id.* at 122.

458. *Id.* at 132 (footnote omitted).

459. Casscells et al., *supra* note 211, at 999 (emphasis added).

460. *Id.* at 999–1001.

461. Cosmides & Tooby, *supra* note 453, at 25.

462. Subjects were told that "1 out of every 1000 Americans has disease X" 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.

463. *Id.* at 26–27.

464. *Id.* at 26–28.

465. *Id.* at 31–33.

probe questions by coloring in the boxes representing the relevant number of individuals, correct answers rose to 92%.[466]

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 Bayesian at all."[467] The authors concluded that as long as their Bayesian instincts are triggered by the expression of probabilities across many tries rather than as a "single-event probability," humans may be "good intuitive statisticians after all."[468] Among subjects with "a concrete *visual* representation of a population that depicts the relevant frequencies, bayesian performance is . . . enhanced to near perfect levels."[469]

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 of 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 Article[470] and "visual explanations" like those Cosmides and Tooby used in their experiments.[471] "Whereas [trial participants] find it hard both to 'believe' and reconstruct the formulaic explanation, they . . . understand the visual explanation."[472]

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

466. *Id.* at 34–37; *see also id.* at 37 ("[P]resent[ing] the problem information as percents and ask[ing] for the answer as a single-event probability, elicited bayesian performance from only 12% of subjects. But by simply translating this problem into frequentist terms we were able to elicit correct bayesian reasoning from 76% of our subjects. By requiring them to create a concrete, visual frequentist representation . . . , we were able to push their performance to 92% correct.").

467. Kahneman & Tversky, *Subjective Probability*, *supra* note 208, at 450.

468. Cosmides & Tooby, *supra* note 453, at 62, 69. For other studies supporting the conclusion that lay decision makers, including jurors, are better able to make sense of statistical information presented as natural frequencies rather than single-event probabilities, see Gerd Gigerenzer et al., *"A 30% Chance of Rain Tomorrow": How Does the Public Understand Probabilistic Weather Forecasts?*, 25 RISK ANALYSIS 623, 625–29 (2005); Jonathan J. Koehler, *On Conveying the Probative Value of DNA Evidence: Frequencies, Likelihood Ratios, and Error Rates*, 67 U. COLO. L. REV. 859, 880–83 (1996); Jonathan J. Koehler & Laura Macchi, *Thinking About Low-Probability Events: An Exemplar-Cuing Theory*, 15 PSYCHOL. SCI. 540 (2004); Koehler, *supra* note 284, at 498–508; Samuel Lindsey et al., *Communicating Statistical DNA Evidence*, 43 JURIMETRICS J. 147, 154–62 (2003).

469. Cosmides & Tooby, *supra* note 453, at 59.

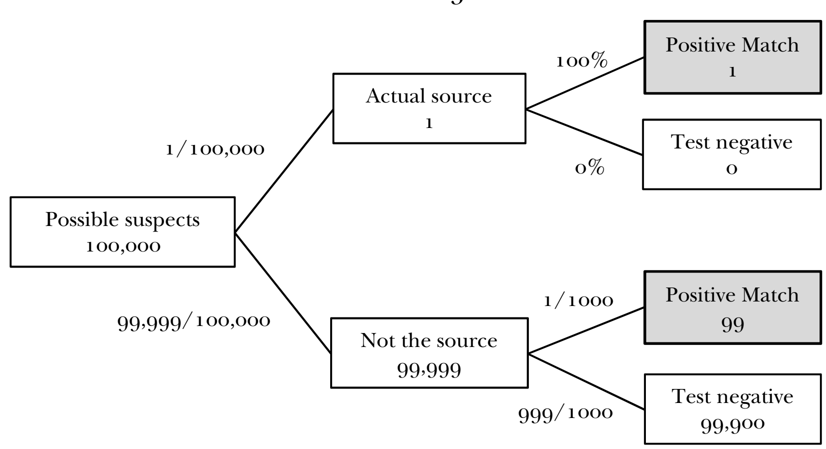
470. *See supra* notes 53–62 and accompanying text.

471. *See* Fenton & Neil, *supra* note 133, at 117–18, 129–35; Fenton & Neil, *"Jury Observation Fallacy," supra* note 453, at 4, 7–11.

472. *See* Fenton & Neil, *supra* note 133, at 132.

evidence proves. The algorithms in both tools are accepted by scientists, yet in the former case, courts do not even require expert testimony to explain the algorithms (even though they sometimes 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 of Bayesian analysis is the willingness of lawyers and jurors to accept the relevant type of calculation because they themselves can perform it.[473] 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. Fenton and Neil used this tool 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 1% of the population. In formulaic terms, the Bayesian equation is *1/99,999 x 1/.001 = 1/99.99 → 1%*. In lieu of a Bayesian formula, Fenton and Neil used an animated version of the decision tree in Figure 3 below 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).[474]

FIGURE 3




*So, 100 have a positive match. But only 1 is guilty.*

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%,[475] Fenton and Neil give jurors an electronic Bayesian

473. *Id.* at 131–34.

474. *Id.* at 133. The authors round the odds to avoid requiring jurors to imagine fractions of people. *Id.*

475. *See supra* notes 179, 449 and accompanying text.

calculator and let them use it to see the effect of different assumptions.[476] For example, jurors might decide there are more suspects (decreasing the prior odds) or that there is 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 "[a]lthough we were able to explain this to you from scratch, there is a standard calculation engine (accepted and validated by the mathematical and statistical community) which will do this calculation . . . instantly for us . . . much like relying on a calculator to do long division."[477] The jury can assume the calculator's accuracy and focus on making good assumptions about the ingredient probabilities.[478]

Fenton and Neil imagine the application of these tools in the *Adams* case discussed above[479] 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,[480] 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 lineup.[481] 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 99.9995% probability of guilt using the frequency the prosecution expert offered) 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 99.9995% to 98.2% using the prosecution's frequency).[482] The jury then decides "if the assumptions in the model are reasonable" and if "the resulting probability of guilt leaves room for [reasonable] doubt."[483]

476. *See* Fenton & Neil, *"Jury Observation Fallacy," supra* note 453, at 12 ("[B]y using Bayesian nets and a tool such as Hugin [a commercially available Bayesian calculator], it is possible to show all of the implications and results of a complex Bayesian argument without requiring any understanding of the underlying . . . mathematics."); HUGIN EXPERT, http://www.hugin.com/ (last visited Oct. 18, 2012) (calling itself "[t]he leading decision support tool").

477. Fenton & Neil, *supra* note 133, at 139.

478. *Id.*

479. *See supra* Part II.B.

480. The defense and prosecution proposed random-match frequencies of, respectively, 1 to 2 million and 1 to 200 million. *See supra* notes 66–68 and accompanying text.

481. Fenton & Neil, *supra* note 133, at 138.

482. *Id.* at 140, 142–43.

483. *Id.* at 142; *see also id.* at 123–24, 133 (applying tools to a criminal case involving a blurry photo showing some but not all of the numbers on a car's license plate and to a medical malpractice case).

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 databases of 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 to remedy the problem, just as the same devices quickly curbed parallel problems with early uses of the multiplication rule in connection with inculpatory DNA.[484] Accordingly, when evidence with ascertainable random-match frequencies is available—which increasingly will be true as the legal market for data-mined frequencies expands[485]—it is no longer fanciful to imagine effective use of Bayesian analysis in court. That analysis could then be used to reveal the aggregate value of non-exclusionary non-matches via tools no more controversial than calculators. As the next Subpart 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 needed 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 Cary Coglianese and David Lazer distinguish three types of regulation.[486] "Technology-based" regulation requires regulated entities to adopt a specified solution to a given problem,[487] 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 over- and under-inclusive when conditions vary and can

484. *See supra* notes 138–66 and accompanying text.

485. *See supra* Part IV.C.

486. Cary Coglianese & David Lazer, *Management-Based Regulation: Prescribing Private Management To Achieve Public Goals*, 37 LAW & SOC'Y REV. 691, 694 (2003).

487. *Id.*

discourage innovation by tying actors to outdated technologies.[488] Alternatively, "performance-based" regulation imposes floors or ceilings on regulated outputs.[489] Limits on the amount of 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. If the desired end state or effective ways to measure it are unknown, however, such regulation falters 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 come to school on the first day with vastly different preparation levels.[490]

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.[491] 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.[492] Planning also can be induced by publicizing undesirable outcomes, such as toxic releases,[493] publicly comparing results attained by similarly situated entities to expose less and more effective operations,[494] and using penalty defaults to threaten entities with severe or unpredictable consequences if they do not reach agreement on solutions with other stakeholders.[495] Management-based strategies work best when the desired outcomes and possible solutions are uncertain or when the conditions faced are highly variable.[496] Both of these prerequisites characterize the criminal justice

488. *Id.* at 701.

489. *Id.*

490. *See id.*

491. *See id.* at 702–03.

492. *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).

493. *See, e.g.*, Charles Sabel et al., *Beyond Backyard Environmentalism: How Communities Are Quietly Refashioning Environmental Regulation*, BOS. REV., Oct./Nov. 1999, *available at* http://bostonreview.net/BR24.5/issue.pdf (discussing federal Toxics Release Inventory and Massachusetts Toxics Use Reduction Act).

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

495. *See, e.g.*, Bradley C. Karkkainen, *Information-Forcing Regulation and Environmental Governance*, *in* LAW AND NEW GOVERNANCE IN THE EU AND THE US 293, 301–05 (Grainne de Burca & Joanne Scott eds., 2007); Charles F. Sabel & William H. Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 HARV. L. REV. 1015, 1067 (2004).

496. *See* Coglianese & Lazer, *supra* note 486, at 704–06.

practices that invite heuristic biases and discourage use of non-exclusionary non-matches.

First, with the recent exception of DNA and a few other categories of forensic evidence,[497] 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.[498] For example, criminal investigators know they can increase the chance of uncovering potentially 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.[499] Additionally, given the speed with which adversarial imperatives have spurred the creation and improvement of technologies 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,[500] it is too soon to mandate particular technologies or performance levels. Instead, state and local customization of solutions to distinct conditions provides a more promising invitation for innovation.

### 2. Regulation of "Big" Evidence of Identity

As Michael Dorf and Katherine Kruse have separately shown, 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" evidence: custodial confessions and eyewitness identifications.[501] 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,[502] and *Wade* hearings enforcing a right to counsel and

---

497. *See, e.g.*, Rhonda M. Wheate & Allan Jamieson, *A Tale of Two Approaches—The NAS Report and the Law Commission Consultation Paper on Forensic Science*, 7 INT'L COMMENT. ON EVIDENCE, Art. 3 (2009) (comparing the United Kingdom's THE FORENSIC SCIENCE REGULATOR, REVIEW OF THE OPTIONS FOR THE ACCREDITATION OF FORENSIC PRACTITIONERS (2009), with NAT'L RESEARCH COUNCIL, *supra* note 96); sources cited *supra* note 133.

498. *See, e.g.*, NAT'L RESEARCH COUNCIL, *supra* note 96, at 14–19 (proposing that an independent federal entity be established to promote the development of forensic sciences); Mnookin et al., *supra* note 96, 27–32; *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 the accuracy of capital verdicts).

499. *See* Daniel O. Larson et al., *Advanced Scientific Methods and Procedures in the Forensic Investigation of Clandestine Graves*, 27 J. CONTEMP. CRIM. JUST. 149, 173–75 (2011) (describing grid and excavation methods forensics teams can use to avoid compromising crime scenes).

500. *See supra* Parts IV.B–C, VI.A.

501. Michael C. Dorf & Barry Friedman, *Shared Constitutional Interpretation*, 2000 SUP. CT. REV. 61, 81–82; Kruse, *supra* note 36, at 648 & n.11, 670–73.

502. Miranda v. Arizona, 384 U.S. 436, 478–79 (1966).

protection against suggestive lineups[503]—neither case actually mandated that procedure.[504] 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 solutions that were equally or more protective and better suited to local conditions. Jurisdictions that adopted alternative measures could avoid the constitutional default rule.[505] As Kruse shows, the Court issued these invitations as much out of a desire for stronger protections than were possible on a one-size-fits-all basis as out of a hope of easing burdens on law enforcement and courts.[506]

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.[507] 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.[508] Recently, however, DNA exonerations of defendants convicted based on false confessions and inaccurate eyewitness identifications have prompted a number of states to begin regulating interrogations and lineups. Much of the regulation has been technology-based, requiring videotaped confessions and sequential double-blind lineups,[509] but Wisconsin and Texas have adopted more comprehensive and flexible management-based solutions.[510]

503. United States v. Wade, 388 U.S. 218, 228–38 (1967).

504. Kruse, *supra* note 36, at 670–73.

505. *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*, 384 U.S. at 490 ("Congress and the States are free to develop their own safeguards for the privilege [against self-incrimination and apply them in lieu of the procedure the court imposed], so long as they are fully as effective as those described above . . . .").

506. Kruse, *supra* note 36, at 671 (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"). The *Miranda* Court explained 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." *Id.* at 672 (quoting *Miranda*, 384 U.S. at 467) (internal quotation marks omitted).

507. *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 36, at 666–72 (discussing other case law).

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

509. *See, e.g.*, OHIO REV. CODE ANN. § 2933.83(B)(1) (West Supp. 2012) (requiring most live and photo lineups to be conducted by officials unaware of which participants are suspects). For a survey of the law regarding electronic recording of custodial interrogations, see generally Gershel, *supra* note 41. *See also* Gudjonsson, *supra* note 106, at 707 (discussing advantages to recording police interrogations); Robert J. Norris et al., *"Than That One Innocent Suffer":*

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.[511] 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 lineup—a factor, ironically, that studies actually correlate with identification *error*.[512] Under the Wisconsin rule, all identifications produced by "unnecessarily suggestive" procedures are inadmissible.[513]

To help local law enforcement agencies cope with *Dubose*'s narrow and unpredictable admissibility standard, the Wisconsin Legislature passed a statute requiring each agency to devise a written policy "to reduce the potential for erroneous identifications."[514] The statute instructs law enforcement agencies (1) to base their initial plans on "model policies" and effective strategies from "other jurisdictions"; (2) to consider including practices shown by research to enhance "objectivity and reliability"; and (3) to revise their plans every two years based on evolving local and statewide experience.[515] For guidance, the Wisconsin Department of Justice has promulgated and periodically updates a Model Policy and Procedure for Eyewitness Identification, citing studies supporting each component.[516] As

---

*Evaluating State Safeguards Against Wrongful Convictions*, 74 ALB. L. REV. 1301, 1304 (2010–2011) (claiming that more than 600 local and county jurisdictions record police interrogations but only 19 states require recording); Andrew D. Rikard, Comment, *Why and How New York Should Enact Mandatory Statewide Eyewitness Identification Procedures*, 73 ALB. L. REV. 1525, 1533–43 (2011) (proposing reforms to New York's eyewitness-identification procedures); Suzanne Smalley, *Police Update Evidence Gathering: Suspect Identification Is Focus of Changes*, BOS. GLOBE (July 20, 2004), http://www.psychology.iastate.edu/~glwells/boston_globe_july_2004.pdf (listing jurisdictions requiring double-blind lineups); sources cited *supra* notes 39–41.

510. H.B. 215, 82d Leg., Reg. Sess., 2011 Tex. Gen. Laws 792 (signed by Gov. Rick Perry, June 17, 2011) (requiring law enforcement agencies to adopt standards and procedures for live and photographic lineups designed by the Law Enforcement Management Institute of Texas or to implement their own equally or more protective policy); *see also infra* notes 514–22 and accompanying text (discussing Wisconsin statutory scheme).

511. State v. Dubose, 2005 WI 126, ¶¶39–41, 285 Wis. 2d 143, 699 N.W.2d 582.

512. *See* Perry v. New Hampshire, 132 S. Ct. 716, 720 (2012); Manson v. Brathwaite, 432 U.S. 98, 106, 114 (1977); Neil v. Biggers, 409 U.S. 188, 199 (1972). *But see* Kruse, *supra* note 36, at 669–70 (criticizing the "reliability" escape hatch for using factors that academic research has associated with inaccuracy as a basis for admitting suggestive identifications).

513. *Dubose*, 2005 WI 126, 919130–33.

514. WIS. STAT. § 175.50(2) (West 2006).

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

516. *See, e.g.*, BUREAU OF TRAINING AND STANDARDS FOR CRIMINAL JUSTICE, WIS. DEP'T OF JUSTICE, MODEL POLICY AND PROCEDURE FOR EYEWITNESS IDENTIFICATION (2010), *available at* http://www.doj.state.wi.us/dles/tns/eyewitnesspublic.pdf ("[E]yewitness evidence is much like trace evidence left at a crime scene. . . . [I]t is susceptible to contamination if not handled properly" (citing Wells & Olson, *supra* note 36, at 286–89)); Kruse, *supra* note 36, at 648 n.10, 687–88. The Wisconsin Model Policy urges investigators to avoid "fishing expeditions" by

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.[517]

Wisconsin took a similar path in regulating confessions. Reacting to the disturbing tactics Milwaukee police used 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 shown to have been infeasible.[518] 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 courts to instruct jurors to "consider the absence of . . . recording" when evaluating custodial statements.[519] 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 and share with other agencies statewide.[520] As Kruse notes, such regimes face a tension between goals of transparency (allowing inter-jurisdictional comparison and sharing) and accountability (creating disincentives to collect and publicize information that may be used against the agency in court).[521] Wisconsin's solution to this tension, which is 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.[522]

---

delaying lineups until alibis and competing leads are investigated and forensic testing is complete. WIS. DEP'T OF JUSTICE, *supra* at 7.

517. Kruse, *supra* note 36, at 685, 689–90.

518. State v. Jerrell C.J. (*In re* Jerrell C.J.), 2005 WI 105, ¶59, 283 Wis. 2d 145, 699 N.W.2d 110.

519. *See* 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); Kruse, *supra* note 36, at 690–93.

520. *See* Kruse, *supra* note 36, at 727–28 ("This new body of information provides rich opportunities for Wisconsin to pursue . . . continuous improvement, public accountability, and cross-jurisdictional learning.").

521. *See id.* at 728, 731.

522. *See, e.g.*, JOSEPH V. REES, HOSTAGES OF EACH OTHER: THE TRANSFORMATION OF NUCLEAR SAFETY SINCE THREE MILE ISLAND 91–120 (1994) (describing the effects of disclosures among CEOs of nuclear-power companies, even absent public disclosures); Sabel et al., *supra*

### 3. Regulation of "Small" Non-Matches

Analogous management-based strategies can 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 given the choice of either adopting other localities' best practices or justifying their own actions to juries. This could be accomplished by allowing the defense to admit evidence of practices that comparable law-enforcement agencies elsewhere in the state use to process evidence of non-exclusionary non-matches and by instructing juries to consider an agency's failure to use effective practices when they evaluate the strength of the evidence.[523] Numerous improvements are available to motivated agencies:

- Comprehensive videotaping of crime scenes with technology for acute magnification and other aids for after-the-fact review, preservation of evidence, and monitoring of investigative rigor;[524]
- Institutional structures or procedures to isolate pretrial investigative, forensic-analysis, and accusatory functions from each other and temper "uniqueness," "irrelevance," and other heuristic mistakes triggered by the early identification of scenarios and suspects;[525]
- 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 suspects;[526]

---

note 493 (describing the regulatory power of public disclosure of comparative toxic-release data).

523. *See supra* notes 514–22 and accompanying text.

524. *See, e.g.*, EDWARD M. ROBINSON, CRIME SCENE PHOTOGRAPHY 582 (2007) (describing the value of photography for crime-scene preservation); Dan Warnock, *The Iraqi Criminal Justice System, an Introduction*, 39 DENV. J. INT'L L. & POL'Y 1, 22 (2010) (listing practices the 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 MAG., Aug.–Sept. 2009, at 39 (similar).

525. *See, e.g.*, Findley & Scott, *supra* note 230, at 393–94 (discussing the value of separating investigative and charging functions to avoid heuristic bias); NAT'L RESEARCH COUNCIL, *supra* note 96, at 24 (recommending that crime laboratories be removed "from the administrative control of law enforcement agencies").

526. *See, e.g.*, 1 MOLLY TREADWAY JOHNSON & LAURAL HOOPER, FED. JUDICIAL CTR., RESOURCE GUIDE FOR MANAGING CAPITAL CASES: FEDERAL DEATH PENALTY TRIALS 10–11 (2004) (describing a multi-tiered process for approving federal death-penalty prosecutions); NAT'L RESEARCH COUNCIL, *supra* note 321, at 84–87 (recommending similar steps to improve intelligence analysis).

- 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 odds;[527]
- Training for the same actors in data-mining techniques for generating frequency information about potentially matching and non-matching clues;[528]
- Open-files discovery of all non-exculpatory non-matches and[529] enforcement of ABA Model Rule 3.8(d), which most states follow but do not apply to prosecutors, obliging state disclosure of all exculpatory evidence;[530]
- 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 by the state as sufficient to establish a reasonable doubt as to guilt;[531]
- Admissibility, subject to the same restriction listed above, 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 that evidence suggests;[532]

527. *See supra* Part VI.A; NAT'L RESEARCH COUNCIL, *supra* note 321, at 84–87 (recommending similar steps to improve intelligence analysis).

528. *See supra* notes 198–200 and accompanying text.

529. *See, e.g.*, Avis E. Buchanan, Op-Ed., *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/gIQATkFMQM_story.html (advocating the open-files discovery already in use in several jurisdictions).

530. *See* ABA Comm. on Ethics and Prof'l Responsibility, Formal Op. 09-454, at 1, 4–5 (2009) (interpreting Rule. 3.8(d) of the *Model Rules of Professional Conduct*, 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 (quoting MODEL RULES OF PROF'L CONDUCT R. 3.8(d) (2009)); Kevin C. McMunigal, *The (Lack of) Enforcement of Prosecutor Disclosure Rules*, 38 HOFSTRA L. REV. 847, 850–55, 860–64 (2010) (discussing states' failure to enforce ABA Opinion 09-454); *see also* Attorney General's Guidelines on Disclosure, CROWN PROSECUTION SERV., http://www.cps.gov.uk/legal/a_to_c/attorney_generals_guidelines_on_disclosure/ (last visited Oct. 18, 2012) (noting that fairness requires disclosure of "all material held by the prosecution that weakens its case or strengthens that of the defence").

531. *See supra* Part V.C.2.

532. *See supra* Part V.C.4.

- Admissibility at trial of expert testimony facilitating the understanding and use of Bayesian analysis and recently developed calculator tools for implementing Bayesian analysis;[533] and
- Broad admissibility of data-mined random-match frequencies, subject to rigorous adversarial testing.[534]

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 other strategies that courts, legislatures, and agencies can adopt to improve the use of non-exclusionary non-matches and other small evidence to 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 recent DNA exonerations have exposed 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. Together, they dupe intuitive decision makers via "uniqueness" and "irrelevance" manifestations of the "certainty" fallacy and other heuristic mistakes into (1) falsely treating eyewitness identifications, confessions, fingerprints, and other "big" evidence as matching a unique trait of the perpetrator and the defendant; and (2) 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—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. What it takes is a little patience; a panoply of old-fashioned adversarial methods of bringing to mind competing uniqueness- and irrelevance-disproving stories; and a multitude of new-fangled methods for mining ubiquitous data for frequency information, making Bayesian analysis accessible to all audiences, and inducing institutions ever-more-effectively to manage risk. Because all of these aids are now at hand, we have all the tools

533. *See supra* Part V.C.5.
534. *See supra* notes 142–66 and accompanying text.

we need to fashion minute probabilistic flecks into treasurably reliable criminal verdicts.