LEGAL PROBABILISM: AN EPISTEMOLOGICAL DISSENT\textsuperscript{1}

Susan Haack

[1] It is clear that some things are almost certain, while others are matters of hazardous conjecture. For a rational man, there is a scale of doubtfulness, from simple logical and arithmetical propositions and perceptive judgments, at one end, to such questions as what language the Myceneans spoke or "what song the Sirens sang" at the other. ... [T]he rational man, who attaches to each proposition the right degree of credibility, will be guided by the mathematical theory of probability when it is applicable. ... The concept "degree of credibility," however, is applicable much more widely than that of mathematical probability.---Bertrand Russell\textsuperscript{2}

Russell’s right. The mathematical calculus of probabilities is perfectly fine in its place; but that place is a limited one. In particular, this mathematical calculus can’t form the backbone of a serious epistemological theory—it sheds almost no light on the crucial concept that Russell calls “rational credibility,” and I call “warrant.” One consequence, as I shall argue here, is that we can’t look to probability theory for an understanding of degrees and standards of proof in the law, but must look, instead, to an older and less formal branch of inquiry: epistemology.

In §1 I argue that legal standards of proof are best understood in terms of the degree to which the evidence presented must warrant the conclusion (of the defendant’s guilt or liability) for a case to be made. In §2 I sketch my analysis of warrant, and show that degrees of warrant cannot be identified with mathematical probabilities. It follows

\textsuperscript{1} © 2011 Susan Haack. All rights reserved. This paper revisits some issues discussed in HAACK forthcoming (a), trying to make some ideas more explicit than I was earlier able to do.

\textsuperscript{2} RUSSELL 1948, p.381 (second emphasis added). ("Perceptive," here, doesn’t have its usual meaning, "insightful," but means, simply, "perceptual.")
immediately that “legal probabilism”—by which I mean the thesis that legal degrees of
proof are to be identified with the probabilities that figure in the mathematical calculus of
probability—is misguided. The rest of the paper will show that legal probabilism fails,
and my approach succeeds, in advancing our understanding of the evidence in two
famous, and famously complicated, cases. I begin, in §3, with Kadane and Schum’s
Bayesian analysis of evidence in the Sacco and Vanzetti case, showing that it is seriously
flawed—and that my approach can do significantly better; and then, to make clear that,
though it isn’t probabilistic, my approach is nonetheless perfectly capable of
accommodating statistical evidence appropriately, in §4 I turn to Finkelstein and Farley’s
Bayesian analysis of the evidence in Collins, showing that it, too, is seriously
flawed—and, again, that my approach can do significantly better.

1. Standards of Proof are Best Understood as Degrees of Warrant
Any legal system needs, somehow, to determine matters of fact; but different legal
systems have (or have had) very different ways of doing this.

In early Anglo-Saxon times, courts relied on trial by oath, in which a defendant
would swear before God that he was not guilty, and “oath-helpers” (also known as
“compurgators” or “con-jurors”\(^3\)) might then be called to swear that his oath was not
foresworn. Lisi Oliver’s book on the history of early English law tells us that around .695
A.D. whether a defendant needed to swear an oath at all, and if so, whether his oath
required the backing of oath-helpers, and if so, how many, depended on his rank: the
word of the king or a bishop was sufficient without his having to swear an oath; a priest
or a deacon had to swear an oath, but didn’t need oath-helpers to back it up; a freeman’s
oath required the support of three oath-helpers of the same rank.\(^4\) Sadakat Kadri’s more

\(^3\) This was presumably the origin of the modern English word “conjuror,” which means
“magician, prestidigitator.”

popular book on the history of trial procedures tells us that in 899 Queen Uta of Germany was acquitted of a charge of adultery only after eighty-two knights swore she was innocent. Any formal defect in the procedure, Kadri continues—even if the witness himself had removed the sacred relic from the reliquary on which he swore!—excused perjury.\textsuperscript{5} Frederick Maitland tells us that the practice of “Wager of Law,” pledging oneself to swear an oath of innocence, wasn’t formally abolished in England until 1833.\textsuperscript{6}

According to Robert Bartlett, a whole variety of legal “ordeal”s emerged during the reign of Charlemagne: trial by cold water, by the cauldron, by the cross, by walking on red-hot ploughshares, etc.\textsuperscript{7} In a “trial by hot iron,” for example, the defendant would be asked to take hold of a red-hot iron bar, and his wound would later be checked to see if it had healed cleanly, or had festered—which was taken to be an indication of guilt.\textsuperscript{8} However, Bartlett continues, trial by ordeal was a last resort, used when there was no other way to discover the truth.\textsuperscript{9} The law of the town of Enns (Austria), granted in the year 1212, provided that, in a case of rape, the accused had the option of trial by ordeal if there were only two witnesses—but not if there were seven or more.\textsuperscript{10} By 1215, when the fourth Lateran Council forbade priests from taking part in such tests, the practice was

\textsuperscript{5} Kadri 2005, pp. 20.

\textsuperscript{6} Maitland 1919, p. 14.

\textsuperscript{7} Bartlett, 1986, pp. 9 ff.

\textsuperscript{8} Id., pp. 21, 33; see also Maitland 1909, Lecture II; Kadri 2005, chapter 1.

\textsuperscript{9} Bartlett 1986, pp. 27-8.

\textsuperscript{10} Id., 29.
already dying out;¹¹ and by 1300 it was, as Bartlett puts it, “everywhere vestigial.”¹²

According to George Neilson, trial by combat—in which the parties to a case would literally fight it out—was introduced in England under William the Conqueror. An Englishman who chose to avoid the duel, Neilson continues, was subject to trial by ordeal; a Norman who chose not to duel, however, had the option of defending himself by oath.¹³ By the time of Henry II, trial by combat had already been confined to a narrow class of cases, and its scope became even narrower as time passed.¹⁴ But English law didn’t officially abolish the practice until 1819, when the Appeal of Murder Act was passed after Abraham Thornton had managed to weasel out of a conviction on a charge of murder by offering to defend himself “with his body.”¹⁵

At its most aggressive, the adversarial culture of U.S. law is sometimes eerily reminiscent of trial by combat; I have occasionally heard scientists of my acquaintance describe their unhappy experience as expert witnesses under cross-examination as “trial by cold water”;¹⁶ and of course we still require witnesses to swear under oath to tell “the truth, the whole truth, and nothing but the truth”—though few, I imagine, now seriously believe that lying under oath will provoke divine punishment. Even today, moreover, in countries that enforce traditional Sharia law, the testimony of a male witness is given

¹¹ MAITLAND 1919, Lecture II suggests that the demise of trial by ordeal was the result of the decision by the fourth Lateran Council; TARUFFO 2009, chapter 1, shows that the Church’s decision merely ratified a shift already taking place in legal practice.

¹² BARTLETT 1986, p.34.

¹³ NEILSON 1890, p.31-2.

¹⁴ Id., pp.33-6.


¹⁶ “Pouring cold water on an idea” is an English idiom for disparaging it.
twice the weight of a woman’s;\textsuperscript{17} and until 2007 the Pakistani law of \textit{zina} required four male, Muslim eye-witnesses to prove a charge of rape.\textsuperscript{18}

Although I will focus here primarily on current U.S. law, and draw on U.S. cases to illustrate my argument, I will be more concerned with degrees of proof than with the distinctively common-law rules of evidence. So at least much of what I say will, I hope, also be relevant to civil-law jurisdictions.

The legal need to determine matters of \textit{fact} leads naturally to rules about burdens and standards of \textit{proof}. “Burden of proof” includes both principles about which party is obliged to produce evidence (also known as the “burden of production”), and principles about which party is obliged to establish the elements of the case to the required degree (also known as the “burden of persuasion”).\textsuperscript{19} Standards of proof specify the degree or level of proof that must be supplied in various kinds of case:\textsuperscript{20} “beyond a reasonable doubt” in a criminal case; “by a preponderance of the evidence” or “more probably than not” in ordinary civil cases; and “clear and convincing” evidence in special circumstances.

\textsuperscript{17} \textsc{Janin and Kahlmeyer 2007}, p. 32.

\textsuperscript{18} Anonymous 2006. On Pakistani law generally, see Sial and Iqbal 2005. (If I understand it correctly, “\textit{zina}” means something like “unlawful sexual conduct,” and includes, e.g., adultery as well as rape.)

\textsuperscript{19} \textsc{Graham 2007}, pp. 577-9.

\textsuperscript{20} Like the common law generally, these standards have gradually evolved over time. One commentator suggests that the highest and most familiar standard, “beyond a reasonable doubt,” has been embedded in Anglo-American law for at least seven hundred years, and perhaps well over a thousand years”(DeLoggio 1986, p. 25). Chadbourn, however, dates the “precise distinction” requiring proof beyond a reasonable doubt in criminal cases to the early 1700s (Chadbourn 1981, vol.9, p. 405). Justice Brennan’s ruling in \textit{Winship} clarifies matters somewhat: “[t]he requirement that guilt of a criminal charge be established by proof beyond a reasonable doubt dates at least from our early years as a nation”; and as McCormick puts it, “[t]he demand for a higher degree of persuasion in criminal cases was recurrently expressed through ancient times, [though] its crystallization into the formula ‘beyond a reasonable doubt’ seems to have occurred as late as 1878.” In \textit{Re Winship}, 379 U.S. 358, 362, 90 S.Ct. 1068, 1071, citing \textsc{McCormick 1954}, pp. 681-2.
such as the termination of parental rights, issues of citizenship, the contents of a last deed, etc.\textsuperscript{21} Then there’s the “reasonable suspicion” or “probable cause” required for a search,\textsuperscript{22} and the requirement in the Texas death-penalty statute that the jury may sentence a defendant to death only if it finds “beyond a reasonable doubt” that “there is a probability” that he will be dangerous in future.\textsuperscript{23}

*Some* assignment of burdens and standards of proof is needed to ensure that a result is reached. And the rationale for the *particular* burdens and standards of proof, likewise, is grounded in policy considerations: most obviously, the requirement that a criminal charge be proven by the prosecution and beyond a reasonable doubt rests on the idea that it is much worse to convict someone of a crime he didn’t commit than to fail to convict someone of a crime he did commit. In *Addington v. Texas* (1979), Chief Justice Burger wrote for the Supreme Court that in a typical civil case involving a monetary dispute, the “preponderance” standard ensures that “[t]he litigants ... share the risk of error in roughly equal fashion”; while in a criminal case, because what the defendant has at stake is so large, we require proof “beyond a reasonable doubt,” to ensure that “our society imposes almost the entire risk of error upon itself.”\textsuperscript{24} Such policy considerations are certainly not beyond the reach of philosophical reflection;\textsuperscript{25} but they are not my

\textsuperscript{21} BROWN, ed. 2006, p. 488.

\textsuperscript{22} The terminology of “reasonable cause” derives from courts’ interpretation of the Fourth Amendment to the U.S. Constitution. Briefly and roughly, there is probable cause “where the known facts and circumstances are sufficient to warrant a man of reasonable prudence in the belief that contraband or evidence of a crime will be found.” *Illinois v. Gates*, 462 U.S. 213, 238 (1983); see also *Ornelas v. United States*, 517 U.S. 690, 696 (1996).

\textsuperscript{23} TEXAS CODE CRIM, PROC. ANN. art.37.071 (West Supp. 2009).


\textsuperscript{25} In this context I think of Percy Bridgman’s reflections on human beings’ tendency to rationalize social institutions that arose in “hit or miss fashion”: “A dog is content to turn around three times before lying down, but a man would have to invent an explanation of it.” BRIDGMAN...
present concern—which is, rather, to understand what these standards of proof amount to.

It's easy enough to order the standards from strongest to weakest: "beyond a reasonable doubt"; "by a preponderance of the evidence"; "clear and convincing evidence"; "reasonable suspicion."\(^{26}\) None of these, however, is very precisely defined; nor, for that matter, is it clear that precise definitions would be desirable even if they were feasible. (Indeed, some federal circuits advise judges that they shouldn't try to define what "reasonable doubt" means, but leave this to the jury to discern.)\(^{27}\) But a more fundamental question really does require an answer: what exactly are degrees of proof degrees of?

In a concurring opinion in *Winship* (1970), Justice Harlan wrote that the function of standards of proof is to "instruct the factfinder concerning the degree of confidence our society thinks he should have in the correctness of factual conclusions for a particular type of adjudication."\(^{28}\) The authors of the 6th edition of a well-known textbook, *McCormick on Evidence* (one of whom, I note, is a leading legal Bayesian, David Kaye), are even more explicitly psychological: the "reasonable doubt" formula, they write, "points to what we are really concerned with, *the state of the jury's mind*"; whereas "preponderance of the evidence" and "clear and convincing evidence" are misleading, because they "divert attention to the evidence." But this, they continue, is a step removed from the essential thing, the degree of the juror's belief; the evidence is only "the

\(^{26}\) Though it may be worth noting that the Appeals Court in *Winship* had suggested that there is only a "tenous difference" between the reasonable doubt and the preponderance standards—a suggestion which, however, the Supreme Court very firmly rejected. *In Re Winship* (note 000 above), 368, 1074.

\(^{27}\) O'Malley et al., eds., 2008, §12:10, p. 164.

\(^{28}\) *In Re. Winship* (note 000 above), 370, 1073 (emphasis mine).
instrument by which the jury’s mind is influenced.”

In my opinion, this has things exactly backwards. Admittedly, the language in which standards of proof are expressed is partly psychological: at any rate, talk of the “burden of persuasion” sounds subjective, suggesting that the attorney’s task is simply to persuade, to induce a certain state of mind in the jurors; and “convincing” also sounds psychological. But the language of standards of proof is also in part epistemological: the “reasonable” in “beyond a reasonable doubt,” like the “clear” in “clear and convincing,” sounds objective, since it apparently refers to the quality of the evidence presented. And this epistemological aspect, I believe, is crucial.

This is not to say that the fact-finder’s degree of confidence in the conclusion is completely irrelevant; after all, a reasonable person will proportion his degree of belief at least approximately to the strength of the evidence—the better the evidence warrants \( p \), the more confidence he will have that \( p \) is true. It *is* to say, however, that the fact-finder’s degree of belief is a distinct, and distinctly secondary, matter; the weight of the evidence is primary.

A sampling of jury instructions on standards of proof confirms this epistemological understanding. I begin with Florida:

- The standard jury instructions in criminal trials provided by the Supreme Court of Florida contrast an “abiding conviction of guilt” with a conviction that “wavers and vacillates”; but are very clear that “it is to the evidence introduced in this trial, and to this alone, that you must look for proof,” and that “[a] reasonable doubt as to the guilt of the defendant may arise from the evidence, conflict in the evidence, or the lack of evidence.”

- Florida instructions on the preponderance standard focus more centrally on degree of persuasion: “[g]reater weight of the evidence’ means the more persuasive and convincing force and effect of the entire evidence in the

---


Florida instructions on “clear and convincing evidence,” however, again focus centrally on the epistemological: “[c]lear and convincing’ evidence is evidence that is precise, explicit, lacking confusion, and of such weight that it produces a firm belief or conviction, without hesitation, about the matter in issue.”

I turn next to Federal guidelines—in which we find a good deal of implicit epistemology:

Federal jury instructions on the criminal standard of proof contrast “reasonable doubt” with merely “possible” doubt: “[p]roof beyond a reasonable doubt ... must be proof of such a convincing character that a reasonable person would not hesitate to rely and act upon it in the most important of his or her affairs.” The first circuit has approved a formulation by Judge Keeton: “reasonable doubt may arise not only from the evidence produced but also from lack of evidence. [I]t exists when, after weighing and considering all the evidence, using reason and common sense, jurors cannot say that they have a settled conviction of the truth of the charge.” The third circuit has approved this formulation: “The doubt must be reasonable. It is not a mere possible or imaginary doubt ... . A reasonable doubt is a fair doubt, based on reason and common sense. ... [A] defendant must never be convicted on mere assumption, conjecture, or speculation.” The sixth circuit adds that a reasonable doubt: “may arise from the evidence, the lack of evidence, or the nature of the evidence”, the eighth circuit writes that the presumption of innocence “can be overcome only if the government proves, beyond a reasonable doubt, each essential element

32 Id.
34 Id., p.165 (citing U.S. v. Cleveland, 106 F.3d 1056, 1062-63 (1st Cir. 1997)).
35 Id., p. 171 (citing U.S. v. Isaac, 134 F.3d 199, 202 (3rd Cir, 1998)).
36 Id., p.174.
of the crime charged”, and the ninth that the verdict “may arise from careful and impartial consideration of all the evidence, or from lack of evidence.”

- Federal jury instructions on the standard of proof in ordinary civil cases, citing model jury instructions for the ninth circuit, explain that, to find that a case has been made by a “preponderance of the evidence,” “you must be persuaded by the evidence that the claim ... is more probably true than not true.”

- And federal jury instructions explain “clear and convincing evidence” as setting a higher standard than “preponderance of the evidence,” requiring that the jury “be persuaded by the evidence that it is highly probable that the claim ... is true.”

None of these, admittedly, is entirely transparent; but (with the possible exception of the Florida instruction on the “preponderance” standard) they all make it very clear that standards of proof should be understood, not as a simple psychological matter of the degree of jurors’ belief, but as primarily an epistemological matter, the degree of belief warranted by the evidence, or the lack of it.

If further confirmation is needed, it can be found by looking at the circumstances in which a judge may direct or overturn a verdict. The Manual of Federal Practice tells us that a court may grant a motion for JMOL (Judgment as a Matter of Law) when either there is a complete absence of proof on one or more material issues, or “there are no controverted issues of fact on which reasonable persons could differ”; i.e., when, viewing the matter in the light most favorable to the party against whom the motion is made,

\[37\] Id., p.187.

\[38\] Id., p.189.


\[40\] Id., §104.03, p. 143.
“there can be only one reasonable conclusion.” JMOL is improper, however, if the evidence is conflicting, or insufficient to make only one conclusion reasonable.\footnote{Givens and Shirey 2010, §7.51, pp. 790-91. “JMOL” covers both directed verdicts (where a judge takes the verdict out of the jury’s hands) and judgments \textit{n.o.v.} or “notwithstanding the verdict” (where a judge overrides a verdict the jury has already brought in).} Federal Rule of Criminal Procedure 29 instructs a judge to direct a verdict of acquittal if “the evidence is insufficient to sustain a conviction.”\footnote{Wright and Henning, 2009, §467, p. 362 (the quotation is from the authors’ description, not the text of the rule).} And the Florida Supreme Court has explained that “Courts should not grant a motion for acquittal unless the evidence is such that a view which the jury may lawfully take of it favorable to the opposite party can be sustained under the law.”\footnote{Allen 2010, §18:13, pp. 750-51 (citing Fitzpatrick v. State, 900 So.2d 495, 507 (Fla. 2005)).}

Moreover, a robustly epistemological understanding of degrees of proof is not only more faithful to the language of jury instructions and the like than a purely psychological understanding; it is also, and more fundamentally, integral to what is required by the role that standards of proof play in legal proceedings. Articulating what that role is, however, requires some subtlety.

In some trials (though not all) the key issue is a factual one: Did Mrs. Coppolino die of natural causes, or was she poisoned?\footnote{Coppolino v. State, 233 So.2d 68 (Fla.2d DCA 1968).} Was it the defendant, Robert Downing, or someone else, who posed as the Reverend Claymore to obtain goods by fraud?\footnote{U.S. v. Downing, 753 F.2d 1224 (3rd Cir. 1985).} Did Ethel Brownstone really sign the document ostensibly giving these valuables to her niece, or is the signature a forgery?\footnote{U.S. v. Starzeczypzel, 880 F.Supp. 1027 (S.D.N.Y. 1995).} Was it Nicola Sacco and Bartolemeo Vanzetti, or the
Morelli gang, who committed the payroll robbery and murder at the Slater and Morrill shoe factory? 47 Was it Janet Collins who knocked over the old lady in an alley and stole her purse, and Malcolm Collins who drove the getaway car, or two other people entirely? 48 Was it his exposure to leaking Toluene that caused Bob Moore’s acute respiratory problems, or his history of heavy smoking and asthma? 49 Was it the Bendectin his mother took while pregnant that caused Jeffrey Blum to be born with club feet, or was this just bad luck? 50 Etc., etc.

In 1996, the Supreme Court wrote in Tehan that “the purpose of a trial is to determine the truth.” 51 And it is true that a verdict is substantively just only if the determination made is factually correct: only, that is, the defendant convicted of the crime really was the perpetrator, the defendant found liable for an injury really was responsible for causing it. 52 Nevertheless, unlike a serious scientific or scholarly inquiry, a trial isn’t much like an open-ended investigation sifting for as long as it takes through all the evidence that can be had; legal determinations of fact are subject both to limitations of time, and to constraints on how evidence may lawfully be obtained and what evidence may lawfully be presented. So what the legal finder of fact is asked to do is not, strictly speaking, to determine whether the defendant is guilty, or is liable, but—eschewing guesswork, whim, prejudice and, as the third circuit puts it, “assumption, conjecture, or

47 See below, §3.

48 See below, §4.

49 Moore v. Ashland Chemical, Inc., 151 F.3d 269 (5th Cir. 1998).


52 See, however, Sindell; Hymovitz.
speculation”—to determine whether the defendant’s guilt or liability has been established to the required degree by the evidence presented.53

This obviously requires standards of proof to be understood, not simply in terms of the fact-finder’s degree of confidence, whether or not it is appropriate given the evidence, but in terms of what it is reasonable to believe in light of the evidence presented. In short, the task of the finder of fact is a paradigmatically epistemological one: “judging of evidence,” in John Stuart Mill’s quaint but exactly appropriate phrase.54 The fact-finder must determine, as Russell would have put it, whether the evidence presented makes the proposition(s) at issue credible to the required degree; or, in my epistemological vocabulary, whether the evidence presented warrants the proposition(s) at issue to the required degree.55

2. Degrees of Warrant aren’t Mathematical Probabilities

As Ian Hacking observes, throughout its history the concept of probability “has [had] two aspects”: it connects both with “the degree of belief warranted by evidence,” and with “the tendency, displayed by some chance devices, to produce stable relative frequencies.”56

53 Since I am speaking primarily about U.S. law, this should, strictly peaking, read “by the admissible evidence presented.” But I won’t keep repeating this; henceforth, it should be understood as implicit.

54 Mill 1843, p.5.

55 For a fuller statement of this, see HAACK 2004 (a). It is, of course, no part of my argument that the legal system always succeeds at this; and to drive that point home I will quote Steve Martini’s fictional attorney Harry Hines, who believes that “most victories in criminal courts are fashioned from the preponderance of perjury. You spin yours and they do theirs, and in the end the side that is most adept at invention wins”; and that “throughout history truth has withered and died of loneliness in most courtrooms.” MARTINI 1994, 420.

56 HACKING 1975, 1.
We commonly use the language of probability or likelihood when we talk about the credibility or warrant of a claim—about how likely is it, given this evidence, that the claim is true, or, unconditionally, about how probable the claim is. I talk this way myself: e.g., when I wonder how probable it is, given the evidence we now have, that there is any causal connection between vaccines and autism;\textsuperscript{57} or how likely it is that Egypt will have a genuinely democratic government five years from now. This is what the *Oxford English Dictionary* takes to be (in British English) the usual sense: “may in view of present evidence be reasonably expected to happen or be the case.”\textsuperscript{58} Similarly, *Merriam-Webster’s Dictionary* gives as its first definition (for American English), “supported by evidence strong enough to establish presumption but not proof.”\textsuperscript{59}

Closely connected with this core usage is our habit of adding “probably,” in conversational speech, when we are reluctant to commit ourselves categorically. When a colleague asks me, “will you be at the talk tomorrow?” and I expect to be, but don’t want to be too firmly committed to making it, my answer may be: “probably.” More generally, we use “probably” and the like when we think something is so, but aren’t sure. When I say that the leftovers in the fridge, though past their best, are “probably” still OK to eat. When the evidence that \( p \) is less than overwhelming, recognizing that \( p \) might turn out to be false, we use “probably” as a way to hedge our commitment.

And of course “probable,” etc., also turn up in the language of legal standards of proof: one formulation of the “preponderance” standard is “more probable than not,” one formulation of “clear and convincing” is “highly probable,” and one formulation of “reasonable suspicion” is “probable cause.” The natural and obvious way to understand these legal uses is in the usual, epistemological sense: how reasonable a claim is in the

\textsuperscript{57} An issue discussed briefly in HAACK 2011.


light of the evidence.

But “probable” and its ilk are also the characteristic idiom of gamblers, statisticians, and actuaries. In these uses, rather than hedging our degree of commitment to a claim, “probable” is part of the content of the claim itself: that there is a 1% chance that a 55 year-old man will die in the next year, a 50% chance that the coin will come up heads, a more than doubled risk of Guillan-Barré Syndrome among those recently vaccinated against swine flu, etc.\textsuperscript{60} This is the idiom regimented by the standard probability calculus, and interpreted in terms of relative frequencies or propensities; and many have hoped to shoehorn legal degrees of proof into this more formal mold. Leibniz, for example, called the theory of probability “natural jurisprudence”,\textsuperscript{61} and George Boole hoped to apply it to “the estimation of the probability of judgments.”\textsuperscript{62} More recently, legal probabilism has been the subject of prolonged debate among practitioners of the New Evidence Scholarship\textsuperscript{63}—a debate that resurfaced in 2010 in the form of an exchange prompted by the recently proposed “restyling” of the Federal Rules of Evidence.\textsuperscript{64} This time, it seems, the trigger was that FRE 401 explains that evidence is relevant just in case it \textit{either raises or lowers the probability} of some fact at issue. It

\textsuperscript{60} My last example alludes to the history leading up to the adoption, by some U.S. courts, of a more-than-doubled-risk criterion for specific causation in toxic tort cases. See HAACK forthcoming (c).

\textsuperscript{61} John Locke, however, had denied that precise rules for calculating legal degrees of proof were possible. I rely on HACKING 1975, pp. 86-7.

\textsuperscript{62} BOOLE 1854, chapter XXI; the quotation is from p. 376. On p. 382, however, he acknowledges that “\textit{from the mere records of the decisions of a court ... it is not possible to deduce any definite conclusion respecting the correctness of the individual judgments of its members}.”

\textsuperscript{63} KAPLAN 1968 seems to have been seminal. See also e.g., LEMPERT 1988; PARKS AND SAKS 2005-6.

\textsuperscript{64} RISINGER et al. 2010 (apparently a lengthy e-mail exchange among the “authors.”). For the proposed “restyling” of the FRE, see ADVISORY COMMITTEE ON EVIDENCE RULES 2009.
should already be clear, however, that we shouldn’t simply assume, just because the word “probable” or “probability” occurs in legal contexts, that we are dealing with mathematical, rather than epistemological, probabilities.

Some critics of legal probabilism have expressed skepticism about the possibility of assigning numbers to degrees of proof. “That ... moral probabilities ... could ever be represented by numbers ... and thus be subject to numerical analysis,” Thomas Starkie wrote in 1842, “cannot but be regarded as visionary and chimerical.” Others have argued that precise calculation of probabilities, even if it were feasible, is usually undesirable: Tribe writes of “The Costs of Precision,” such as “the dwarfting of soft variables,” and suggests that, in criminal trials, a Bayesian approach may encourage a “presumption of guilt.” I agree: it isn’t feasible to put precise numbers on degrees of proof, nor would it necessarily be desirable to do so even if we could. But my objection to legal probabilism is even more fundamental: I think identifying legal degrees of proof, i.e., degrees of epistemic warrant, with mathematical probabilities is a kind of equivocation. On this, I’m with Richard von Mises, who long ago averred that “probability theory has nothing to do with such questions as: Is there a probability of Germany being at some time in the future at war with Liberia?”

There will be some overlap between my arguments and those of other critics of legal probabilism, such as L. Jonathan Cohen (on the philosophical side) and Leonard


66 Tribe 1971, pp.1358 (the costs of precision), 1361 (the dwarfting of soft variables), and1368 (presumption of guilt). (This paper was in part a response to Finkelstein and Farley 1969-70, discussed at length in §4 below.)


Jaffee and Ronald Allen (on the legal side). I don’t, however, endorse Cohen’s reconstrual of probabilities or degrees of proof in terms of a quasi-Baconian “inductive logic”; for I believe the moral of the “grue” paradox is that there can be no formal inductive logic. And while there is some distant affinity between the concept of explanatory integration that will play a role in my epistemological account and the “explanatory stories” that play a role in Allen’s proposed “reconceptualization of civil trials,” these two concepts are different, as is their role in his and my overall accounts—and, indeed, the purposes of the accounts themselves.

My arguments will be specifically epistemological in character: that the epistemological and the statistical uses of “probable” are importantly different; and moreover that epistemological probabilities—i.e., degrees of rational credibility or warrant—don’t fit the profile of the Kolmogorov axioms of the standard probability calculus. So, obviously, I need to begin with at least a sketch of the main contours of my understanding of warrant.

Briefly and roughly: how warranted a claim is depends on how good the evidence

---

69 JAFFEE 1984-5.

70 ALLEN 1986.

71 For the “grue” paradox, see GOODMAN, 1954. On its consequences for the possibility of a syntactically-characterizable inductive logic, see HAACK 2003, pp. 84-6. I also note that Prof. Allen argues that Cohen’s account runs into the very same problems as the more conventional standard-probabilist approach he intends it to replace. ALLEN 1986, pp. 415 ff.

72 ALLEN 1986, § III. See also LEMPERT 1988, pp. 84 ff. for a very clear summary of Allen’s proposal. As his title indicates, Allen’s approach is intended to apply only to civil trials—which itself reveals that his preoccupation are more legal than, like mine, epistemological. I note for the record that my epistemological account was developed long before I read either Cohen or Allen.

73 Of course, this can be only a sketch. For the full story, see HAACK 1993/2009, chapter 4 and HAACK 2003 chapter 3.
is. A little more precisely: evidence includes both experiential evidence and reasons, and these work together like clues and already-completed entries in a crossword puzzle. Crucially important for present purposes, reasons ramify, like crossword entries; and what makes the evidence with respect to a claim better or worse is analogous to what makes a crossword entry more or less reasonable:

(i) How supportive the evidence is; analogue: how well a crossword entry fits with the clue and intersecting completed entries. Evidence may be supportive (positive, favorable), undermining (negative, unfavorable), or neutral (irrelevant) with respect to some conclusion.

(ii) How secure the reasons are, independent of the claim in question; analogue: how reasonable the competed intersecting entries are, independent of the entry in question. The better the independent security of positive reasons, the more warranted the conclusion, but the better the independent security of negative reasons, the less warranted the conclusion.

(iii) How comprehensive the evidence is, i.e., how much of the relevant evidence it includes; analogue: how much of the crossword has been completed. More comprehensive evidence gives more warrant to a conclusion than less comprehensive evidence does iff the additional evidence is at least as favorable as the rest.

Briefly and roughly, how supportive evidence is depends on the contribution it makes to the explanatory integration of evidence-plus-conclusion—or, even more briefly and even more roughly, on how well evidence and conclusion fit together in an explanatory account. In line with this, whether and if so how much support a particular piece of evidence gives to a conclusion depends on whether and if so how much adding that piece of evidence contributes to the explanatory integration of the whole. Think of a

---

74 The exact character and role of experiential evidence raises challenging philosophical questions; but in the present context these can be set aside as only marginally relevant.

75 The idea of comprehensiveness seems to be implicit in the sixth circuit’s instruction that reasonable doubt may arise not only from the evidence, but from the lack of evidence.
crossword entry that is well-supported by its clue and other entries, but where one of those other entries in particular—perhaps the one with an “x” in the middle—is especially significant because, without it, the overall fit would be much looser: this is the analogue of a piece of evidence the support of which is particularly crucial with respect to some conclusion. It may be worth adding that, for this reason, my approach is neither atomistic nor fully holistic; it is a kind of articulated holism.

This theory is foundherentist: by which I mean that, like foundationalism but unlike coherentism, it acknowledges the role of experiential evidence; while at the same time, like coherentism but unlike foundationalism, it acknowledges the pervasive relations of mutual support among our beliefs. It is also worldly: by which I mean that it recognizes that the determinants of evidential quality, and hence of warrant, are not entirely formal, but in part material. Whether evidence is relevant to a claim may depend on facts about the world: e.g., whether the defendant’s wound’s having healed cleanly is relevant to his guilt, as the practice of trial by hot iron presupposed, depends on whether it is true that God will protect the falsely accused; whether the effect of this drug on rats is relevant to its effect on humans, as the catastrophically misleading animal studies of the effects of Thalidomide presupposed, depends on whether it is true that rats are similar to humans in the salient physiological respects. And so on.

As this reveals, the explanation of relevant evidence found in the FRE (whether in the old or the “restyled” version), as evidence that either raises or lowers the probability

---

76 Foundationalism and coherentism are the two traditionally rival epistemological theories in the field; but, as I have shown, they don’t exhaust the possibilities. For details, see Haack 1993/2009 chapter 1; for a more summary treatment, Haack 1998.

77 All this is spelled out in detail in Haack 2003, pp. 76-7

78 The tragedy of Thalidomide could have been avoided had the scientists concerned realized that, in the salient respects, the rats weren’t like humans: the drug didn’t send them to sleep, because they weren’t absorbing it. Stephen and Brynner 2001.
of some fact at issue, leaves something to be desired. Evidence is relevant iff it affects the degree of supportiveness of the evidence overall; i.e., iff adding this piece of evidence either contributes to or detracts from the explanatory integration of evidence-plus-conclusion—which, as we just saw, may depend on facts about the world. So the legal definition is potentially misleading in two ways: overtly, by suggesting that relevance is a matter of whether this evidence changes the mathematical probability of a some fact at issue, rather than whether it changes its epistemic status; and covertly, by its silence about the material character of relevance.

It should already be apparent that, if the concepts of quality of evidence and degree of warrant are as complex, as subtle, as multi-dimensional, and as worldly as this account suggests, the mathematical theory of probabilities couldn’t possibly, by itself, constitute a theory of warrant. This isn’t yet enough to show that, if we had an adequate theory of warrant in hand, the calculus of probabilities couldn’t then serve as a way of computing degrees of warrant. But I don’t believe the probability calculus can do this, either, for at least three reasons:

- since quality of evidence has several distinct dimensions (supportiveness, independent security, comprehensiveness), and there is no way to rank relative success and failure across these different factors, there is no guarantee even of a linear ordering of degrees of warrant;\(^79\)

- while the probability of \(p\) and the probability of not-\(p\) must add up to 1, when there is no evidence, or only very weak evidence, either way, neither \(p\) nor not-\(p\) may be warranted to any degree; and

- while the probability of \(p\) and \(q\) (for independent \(p\) and \(q\)) is the product of the two, and hence, unless both are 1, less than the probability of either, the warrant of a conjunction may be higher than the warrant of its

\(^79\) A point also made in Keynes 1921, pp. 27-8.
The second and third points have a clear bearing on issues about legal proof.

The “preponderance of the evidence” standard can’t be adequately understood in terms simply of which party produces more evidence; it’s not a matter of counting the number of witnesses proffered or hefting the weight of the documents presented by each side. But the alternative form of words, “more probable than not,” requires careful handling. It’s not sufficient that one party produce somewhat better evidence than the other; what’s required is that the party with the burden of proof produce evidence good enough to warrant the conclusion to the required degree. So, exactly as my second point would lead one to expect, even the preponderance standard—which, in virtue of the formulation “more probable than not,” sounds the most amenable to a probabilist interpretation—in fact resists it.

With respect to the third point, it is worth noting, first of all, that if legal degrees of prof were mathematical probabilities, jury instructions to the effect that each element of the case should be established to the required degree would fall short of the mark; as is obvious when you realize that—supposing for the sake of argument (though only for the sake of argument) that proof beyond a reasonable doubt requires a mathematical probability of 0.98—evidence establishing each of three independent elements to this degree would have a joint probability of only 0.85, well below the threshold (and evidence establishing each of three independent elements to a probability of 0.51 would have a joint probability of only 0.13).

---

80 The negative point, that the application of conventional probability theory here runs into problems with negation and conjunction, was made by Cohen. But, as will, I hope, be obvious, my positive point, that an adequate account of warrant yields the right results on both issues, was not.

81 Richard Posner reports that federal judges’ estimates of what degree of probability “beyond a reasonable doubt” represents range from 75% to 95%. POSNER 2001, p.367.
In fact, a congeries of evidence can warrant a conclusion to a _higher_ degree than any of its components alone would do. I have made this argument elsewhere, in the context of legal disputes in toxic-tort cases about so-called "weight of evidence methodology."\(^{82}\) Briefly and roughly: combining evidence will raise the degree of warrant of a conclusion just in case it makes the evidence overall more supportive of the conclusion; and/or improves the independent security of component pieces of evidence; and/or introduces new evidence which is no less favorable to the conclusion than the more restricted evidence. This won’t invariably happen: the point isn’t that combined evidence is _always_ stronger than any of its components, but that it _sometimes_ is. For example, epidemiological evidence of increased risk of D among those exposed to S may, or may not, interlock with other studies. If we also know that the same substance, S, in comparable doses, gives rise the same kind of damage, D, in animals known to be physiologically similar to humans in the relevant respects, and that toxicology shows that chemical compounds closely similar to S are also associated with D, then the combined evidence may give us better reason to think that S causes D than any of these pieces of evidence alone. Similarly, evidence that the defendant had a strong motive to want the victim dead may interlock with evidence that he owned a gun of the right caliber, and with evidence that he was seen fleeing the scene of the crime, to give more warrant to the conclusion that he did it than any of these pieces of evidence alone.

If, as I have argued, legal degrees of proof are best construed as degrees of rational credibility or warrant and if, as I have also argued, degrees of rational credibility or

---

\(^{82}\) See, e.g., _Oxendine v. Merrell Dow Pharmaceuticals_, 506 A.2d 1100, 1108 (D.C 1986) ("[Dr. Done] conceded his inability to conclude that Bendectin is a teratogen based on any one of the individual studies which he discussed, but he also made quite clear that all these studies must be viewed together, and that, so viewed, they supported his conclusion"). _Gen. Elec. Co. v. Joiner_, 522 U.S. 136, 153, 118 S.Ct. 512, 522 (1997) (Justice Stevens, dissenting) ("It is not intrinsically 'unscientific' for experienced professionals to arrive at a conclusion by weighing all available evidence."). The argument sketched here is given in full detail in HAACK 2008.
warrant cannot be identified with mathematical probabilities, legal probabilism is misguided. Still, as we say in English, “the proof of the pudding is in the eating.”\(^{83}\) Simply showing that legal probabilism is misguided doesn’t, by itself, enable us to get a better grip on complex evidence. So the next move is to put my epistemological approach to work, illustrating how it helps us understand degrees of proof by showing its advantages over avowedly probabilistic approaches to the evidence in two famous, or perhaps notorious, cases: *Commonwealth v. Sacco and Vanzetti*,\(^{84}\) and *People v. Collins*.\(^{85}\)


Champions of what is perhaps the most challenging form of legal probabilism, subjective Bayesianism,\(^{86}\) can be expected to object that my contrast between probabilistic and epistemological approaches is false. Their probabilistic approach, they will insist, is epistemological. For on this approach, they will argue, probabilities are understood subjectively, as degrees of belief; the degrees of belief a person gives various propositions are rational if they are consistent; and they are consistent if they satisfy the axioms of the calculus of probabilities. Bayes’s Theorem then provides a way to adjust

---

\(^{83}\) As this proverb reveals, the word “proof”—which has the same root as “probable”—originally meant “test.”

\(^{84}\) For transcripts of the trial and subsequent proceedings see ANONYMOUS 1969.

\(^{85}\) *People v. Collins*, 438 P.2d 33 (1968).

\(^{86}\) The history here is a bit tangled. Bayes’s Theorem is a uncontroversial theorem of the probability calculus (named after the Rev. Thomas Bayes, who proved it). Bayes himself, however, didn’t publish the proof, which appeared in the *Philosophical Transactions of the Royal Academy*, after his death, in 1763, thanks to his executor Richard Price; and it was Price, and not Bayes, who suggested that the theorem might have broader implications. SCHUM 1994, 48-9. What is now called “Subjective Bayesianism” involves a specific interpretation of probability (as degrees of belief) and a specific use of Bayes’s Theorem (as a means of adjusting one’s degrees of belief in light of new evidence); and is, well, *not* uncontroversial.
(“update”) one’s prior degrees of belief, or prior probabilities, as new evidence comes in. So, the subjective Bayesian concludes—since what legal standards of proof demand is, precisely, that the fact-finder have a certain degree of rational belief in the conclusion, and since degree of belief under probabilistic coherence constraints is degree of rational belief—legal degrees of proof can be identified with degrees of rational belief; understood as he proposes.

This is the theoretical background against which Jay Kadane and David Schum offer their analysis of the evidence presented in the trial of Nicola Sacco and Bartolemeo Vanzetti,87 two Italian immigrants accused of a murder committed during a payroll robbery in South Braintree, Massachusetts in 1920. The central issue at trial was a factual one—whether Sacco and Vanzetti really were the perpetrators. The evidence presented was complicated to say the least; a hundred and fifty-eight witnesses testified—ninety-nine for the prosecution, and fifty-nine for the defense: about Sacco and Vanzetti’s having been seen at the scene of the crime, about the reliability or otherwise of the alibis they gave, about the likelihood that the fatal bullet was fired from Sacco’s gun, about a cap allegedly belonging to Sacco found at the scene, about the get-away car, about the defendants’ subsequent actions—allegedly revealing consciousness of guilt—and so forth and so on. The jury found Sacco and Vanzetti guilty, and they were sentenced to death. After numerous motions and appeals—all of them, including one based on Celestino Madeiros’s confession that it wasn’t Sacco and Vanzetti, but he and other members of the Morelli gang who committed the crime, unsuccessful—Sacco and Vanzetti were executed in August 1927.88

Sacco and Vanzetti were anarchists; and at the time of the trial, after the Russian

87 KADANE AND SCHUM 1996.

88 For a brief history, see TOPP 2005, “Introduction,” pp. 1-51, and “Chronology of Events Related to the Sacco and Vanzetti Case,” pp. 185-8. Madeiros was executed the same day, for a different murder. Id., p. 188.
revolution, the U.S. was in the grip of a “Red Scare.” In 1918 Congress had passed the Sedition Act, making it a federal offense, when the country was at war, to “wilfully utter, print, write, or publish any disloyal, profane, scurrilous or abusive language about the form of government of the United States ... ”,89 and many thought Sacco and Vanzetti had been scapegoated for their political views. In 1927, the year they were executed, Felix Frankfurter published a short book pointing out many flaws both in the evidence on which they were convicted and in the judge’s rulings;90 on the day of the execution, Upton Sinclair began writing Boston, a passionate novel based on the case;91 the following year, Edna St. Vincent Millay wrote a poem, “Justice Denied in Massachusetts,” on behalf of the two men;92 on the fiftieth anniversary of their execution, the then-governor of Massachusetts, Michael Dukakis, issued a proclamation acknowledging that “the atmosphere of [Sacco and Vanzetti’s] trials and appeals was permeated by prejudice against foreigners and hostility toward unorthodox political views,” and declaring August 23, 1977, “Nicola Sacco and Bartolomeo Vanzetti Memorial Day.”93 The case continues to excite controversy to this day. So, if Kadane and Schum have really succeeded in shedding light on the evidence, this would be of more than theoretical significance.

In fact, however, Kadane and Schum’s analysis is more confusing than illuminating. For one thing, their title, promising a “probabilistic analysis” of the Sacco and Vanzetti evidence, is misleading. While they do offer a breakdown of the evidence


90 FRANKFURTER 1927.

91 SINCLAIR 1928.

92 MILLAY 1928.

93 This Proclamation is reprinted in TOPP 2005, pp. 182-4.
into component parts (represented by means of Wigmore diagrams), they do this, so to speak, entirely by hand; he role played by probability theory is, rather, in the synthesis of the evidence to draw conclusions. For another, though Kadane and Schum are well aware of Cohen’s objections to construing degrees of proof as standard mathematical probabilities, they offer no reply; instead, after noting that the Federal Rules define relevant evidence as evidence that either raises or lowers the probability of some fact at issue, they simply take for granted that legal degrees of proof are mathematical probabilities. Moreover, they offer no categorical conclusions, but only various posterior probabilities that Sacco was involved in the crime, or that Vanzetti was, given various assignments of prior probabilities to various items of evidence, and various assignments of conditional probabilities. After hundreds of pages of diagrams and calculations, this is disappointing, to say the least. But there are far worse problems.

First: the fact that Kadane and Schum offer a whole range of very different conclusions, all of them probabilistically consistent, reveals that probabilistic consistency is not sufficient to guarantee rational or reasonable degrees of belief—which would

94 WIGMORE 1913; for a summary account, see GODWIN 2000.

95 An earlier book of Schum’s had alluded to such concepts as missing evidence, relevance, and credibility, but so far as I can determine didn’t come even close to developing a real theory of warrant. SCHUM 1994, pp. 15, 29 (on missing evidence), 69 ff. (on relevance), and 207 (on credibility).

96 KADANE AND SCHUM 1996, pp. 239-40. Schum had discussed Cohen’s approach in some detail in earlier work (SCHUM 1979; SCHUM 1994); but I wasn’t able to identify replies to Cohen’s objections there, either, only a suggestions, in the earlier piece, that the benefits of Bayesianism are such that the “paradoxes” of negation, conjunction, etc., can be ingored.

97 Id., p. 50. SCHUM 1994 maintains that people’s judgments of relevance depend on their “standpoint”; but sometimes seems to elide this into the very different claim that relevance itself is standpoint-relative (pp. 71-3).

98 Qua foundherentist, of course, I don’t believe any kind of coherence, on its own, is sufficient for warrant. See HAACK 1993/2009 chapter 3; 2004 (b).
surely also require, if we are to speak in these terms at all, reasonable prior degrees of belief and reasonable conditional degrees of belief.

Second: Kadane and Schum acknowledge that the probabilities in which they traffic can be understood neither in terms of the doctrine of chances, nor in statistical terms; rather, they say, they are “personal, subjective, or [i.e., i.e.] epistemic probabilities.” But “personal” doesn’t mean the same as “subjective,” and neither means the same as “epistemic”; and their verbal fudging here is a symptom of a deeper issue. Suppose for the sake of argument (though only for the sake of argument) that the probabilities they calculate are degrees of belief: whose degrees of belief are they? Their own, Kadane and Schum reply, and some other scholars’. As this reveals, Kadane and Schum are piggy-backing on the epistemic judgment of people who have studied the case. The supposed identification of probabilities with these people’s (subjective) degrees of belief is spurious; what we’re really dealing with are experts’ opinions about (objective) epistemic likelihoods.

And so, third: Kadane and Schum’s many pages of calculations may create the impression that the calculus of probabilities is doing some serious epistemological heavy lifting; but this is an illusion. The only epistemology going on here is Kadane and Schum’s entirely informal decomposition of the evidence into component elements, and their and other experts’ entirely informal appraisals of the worth of the evidence. The mathematics, when it isn’t downright misleading, is mostly decorative.

To show that my account does better—rather than provide the book-length treatment looking at all the evidence would require—I will take my cues from Frankfurter’s discussion of the various elements of the evidence in the case. As I said

99 KADANE AND SCHUM 1996, pp. 24, 120, 159. They even write that “[w]hat nonindependence means is that knowledge of one item of evidence may influence our judgment of the probative force of another.” Id., p. 129 (italics mine).

100 Id., pp. 239-40.
before, the idiom of probability has both epistemological and mathematical/statistical uses; and when Frankfurter writes that “[e]very reasonable probability points away from Sacco and Vanzetti [and] every reasonable probability points toward the Morelli gang,”\textsuperscript{101} it’s quite clear that it is the epistemological use he intends. Moreover, as we will see, his epistemological observations about the weaknesses in the evidence have—if you’ll pardon the anachronism\textsuperscript{102}—a decidedly foundherentist cast.

Frankfurter points out, for example, the numerous flaws in the eye-witness testimony placing Sacco at the scene. At trial, a year after the robbery and murder took place, Mary Splaine testified with great confidence that she had seen Sacco—whom she described in great (though not entirely accurate) detail—in the getaway car. She had only seen the man she identified as Sacco—who was allegedly in a getaway car traveling at 15 to 18 miles an hour—from a distance of 60 to 80 feet, and for only one-and-a-half to three seconds. She laid great stress on the fact that the man she saw had a “good-sized left hand, a hand that denoted strength”; but Sacco’s hands were smaller than average. Moreover, she had earlier told police that she couldn’t identify the person she saw, and at one point she had identified a different man, who, however, turned out to have been in jail at the time of the crime; and she picked out Sacco only after she had seen him several times at the police station and in court.\textsuperscript{103} Another eye-witness, Frances Devlin, who also claimed to have seen Sacco in the car, had also said earlier that she couldn’t positively identify him.\textsuperscript{104} Louis Pelzer said he had seen Sacco too; but fellow-workmen testified

\textsuperscript{101} Frankfurter 1927, p.101.

\textsuperscript{102} Only the term is anachronistic, I should add: for my foundherentism was from the beginning intended in part as an articulation of the standards of better and worse evidence, more and less warranted belief, implicit in our everyday assessments of evidence. See especially Haack 1993/2009 chapter 1. Molière?

\textsuperscript{103} Id., pp. 11-15; the quotation is from p. 11.

\textsuperscript{104} Id., p. 15.
that he had been hiding under a bench at the time of the shooting.\textsuperscript{105} Two other eyewitnesses, Ferguson and Pierce, who saw the crime from one floor above where Splaine and Devlin were standing, found it impossible to make any identification at all.\textsuperscript{106} Etc., etc. In short, as I would say:

- though, if it were true that Splaine, Pelzer and Devlin saw Sacco in the getaway car, this would strongly support the proposition that Sacco was involved, this eyewitness testimony was, to say the least, sadly lacking in independent security. There is a far better explanation of why Splaine identified Sacco than that he was in fact the man she saw at the crime scene—that she had seen him at the police station and in court; there is a far better explanation of why Pelzer identified Sacco than that he was the man he saw at the crime scene—that he didn’t want to admit he had been hiding after he heard shots; and it seems likely that Devlin, who later insisted she had never had any doubt that Sacco was the man she saw, had grown more certain over time, perhaps because of “the immensity of the crime and everything.”\textsuperscript{107}

Frankfurter also looks in detail at the firearms testimony. Expert witness Captain William H. Proctor testified at trial that “bullet 3” was “consistent with its having been fired from Sacco’s pistol,\textsuperscript{108} and Judge Thayer interpreted this to mean that “it was [Sacco’s] pistol that fired the shot.” However, not all of Proctor’s evidence was given at trial; in a subsequent affidavit he said that “[a]t no time was [he] able to find any evidence whatever which convinced [him] that [this bullet] came from Sacco’s pistol.”\textsuperscript{109} “By prearrangement,” Frankfurter comments, “the prosecution brought before the jury a piece of evidence apparently most damaging to the defendants, when in fact the full truth

\textsuperscript{105} Id., pp. 17-18.

\textsuperscript{106} Id., p. 16.

\textsuperscript{107} Id., p.15.

\textsuperscript{108} Five other bullets were also found in the dead bodies at the scene; but the evidence excluded the possibility that any of these others had been fired by Sacco or by Vanzetti. Id., p.76.

\textsuperscript{109} Id., pp. 76-9.
concerning this evidence was very favorable to them.\textsuperscript{110} In short, as I would say:

- \textit{the firearms testimony presented at trial was sadly lacking in comprehensiveness, so that the conclusion that one of the fatal bullets came from Sacco's gun was very poorly warranted.}

The lies Sacco and Vanzetti told at the police station were presented by the prosecution as indicating "consciousness of guilt"; but given that this was a period of wholesale arrests and deportations of aliens suspected of Communist sympathies, Sacco and Vanzetti may have lied, not because they were guilty of the Braintree crime, but for fear they were in trouble over their political radicalism.\textsuperscript{111} Moreover, Frankfurter points out, none of the stolen $16,000 was ever found in Sacco or Vanzetti's possession; and neither man went into hiding or left the country after the crime—both stayed in their old lodgings, and both continued to pursue their old work.\textsuperscript{112} However, he continues, shortly after the crime Madeiros mysteriously came into possession of $2,800—just the kind of sum one might expect to have been his share of the loot.\textsuperscript{113} In short, as I would say:

- \textit{the lies Sacco and Vanzetti told the police offer only very weak support to the conclusion that they were trying to hide their complicity in the Braintree crime; for the explanatory integration of this evidence with that conclusion is no better than its integration with the conclusion that they were trying to avoid punishment for their political views. Moreover, the additional evidence about their subsequent behavior gives further support to the latter conclusion, while undermining the former.}

I could go on; but this is enough, I hope, to illustrate the advantages of the foundherentist approach over Kadane and Schum's impressively complicated, but ultimately unilluminating, "probabilistic analysis" of this evidence.

\textsuperscript{110} \textit{Id.}, p. 76.

\textsuperscript{111} \textit{Id.}, pp. 35 ff.

\textsuperscript{112} \textit{Id.}, pp. 35-6.

\textsuperscript{113} \textit{Id.}, p. 114.
Well, yes, you may be thinking: but even if the foundherentist epistemological approach is, as you say, better in this instance, this really isn’t the kind of case that best illustrates the virtues of legal probabilism. After all, the Sacco and Vanzetti case involved no statistical evidence; but this, surely, is where probabilism comes into its own. I disagree; but to show why, I need to turn to my second illustration.

4. People v. Collins: Epistemology Trumps Probability Again
There’s no question that statistical evidence—from DNA analyses in criminal cases to epidemiological studies in toxic-tort cases, actuarial calculations in wrongful-death suits, etc., etc.—plays a very significant role in the law; and quite properly so. And there is no question, either, that the mathematical calculus of probabilities is applicable to statistical evidence. It doesn’t follow, however, that the calculus of probabilities can illuminate the role statistical evidence plays in the context of the larger body of evidence in a case; and, as we shall soon see, it isn’t true. On the contrary, in fact: by tempting us to confuse statistical probabilities with degrees of proof, legal probabilism can seduce us into forgetting that the statistical evidence in a case should be treated as one piece of evidence among many.

Here I am deliberately echoing the words of the Supreme Court of New Jersey in Landrigan, rejecting the idea that epidemiological evidence of a doubling of risk is sufficient to establish specific causation in a toxic-tort case: “a relative risk of 2.0 is not so much a password to a finding of causation as one piece of evidence among many.” This gets the key epistemological point right. And I am also deliberately setting myself against the words of the article on epidemiology in the first edition of the Federal Reference Manual on Scientific Evidence, a couple of years earlier: “[t]he relative risk from an epidemiological study can be adapted to [the civil] 50% plus standard to yield a

probability or likelihood than an agent caused an individual’s disease.”115 This commits precisely the confusion against which I am warning.

Though, as the previous paragraph indicates, my understanding of degrees of proof is intended to apply quite generally, I shall again focus here on the evidence in a criminal case. *People v. Collins*, however, was hardly, like the Sacco and Vanzetti case, a *cause célèbre*; on the contrary, it was the kind of case only an evidence scholar could love—frankly piddling, except for a bizarre epistemological twist. This twist was what led Finkelstein and Farley to open what would become a key paper in the legal-probabilist literature, offering a “Bayesian approach to identification evidence,” with a commentary on *Collins*.116 But as we shall see, though Finkelstein and Farley’s effort is a bit more philosophically sophisticated than Kadane and Schum’s, it is seriously flawed nonetheless.

But let me begin at the beginning: with the old lady who was knocked down in an alley and had her purse stolen. The victim described the robber as a young woman weighing roughly 145 pounds, with light blond hair in a ponytail; and another eyewitness said he had seen a blond woman run out of the alley, and jump into a yellow car driven by a black man with a mustache and a beard. Janet and Malcolm Collins, who were accused of the crime, were a mixed-race couple, and drove an elderly yellow car.117 However, the victim couldn’t positively identify Janet, and the other eyewitness’s identification of

115 *Bailey* et al. 1990, 168. (The article on epidemiology in the second, 2000 edition of the *Reference Manual* makes a very similar claim, but backs it up by reference to a string of cases.) *Green* et al 2000, p. 384. For a detailed critique of the “doubling of risk” criterion, see *Haack* forthcoming (c).

116 *Finkelstein* and *Farley* 1969-70.

117 Officer Kinsey, while driving home from work, saw the defendants in their yellow Lincoln, placed them under surveillance, and followed them home. The couple didn’t, by the way, match the eyewitnesses’ description exactly: Janet’s hair was dark blond, and Malcolm didn’t have a beard. *Collins* (note 000 above), 322.
Malcolm was shaky. So, to shore up a not-very-strong case, at trial the prosecution first introduced a mathematics instructor to testify—on the basis purely of a statistical argument itself based, not on evidence, but on sheer assumption—that, given the “product rule,” the odds against there being another such couple around (blond woman, black man with facial hair, with a yellow car) were an astronomical 1 in 12 million; and then told the jury that these made-up numbers were “conservative” estimates, and that they had been given “mathematical proof” that the Collineses were guilty.\textsuperscript{118} The jury duly convicted.

Janet served her time; but Malcolm appealed, and in 1968 was granted a new trial.\textsuperscript{119} Justice Sullivan’s argument, in the ruling granting the new trial, was two-pronged: first, the statistics that apparently swayed the jury were invented, utterly devoid of any factual basis; second, even if they had been real, “no mathematical formula could ever establish ... that the prosecution’s eyewitnesses correctly observed and accurately described the distinctive features ... linking the defendants to the crime.”\textsuperscript{120} The first point is undeniable; the second—while also, I believe, quite correct—raises questions about what role, exactly, statistical evidence plays in the identification of a specific perpetrator or perpetrators.

Statistical identification evidence, Finkelstein and Farley agree, shouldn’t normally be sufficient, but needs to be accompanied by other evidence forming the basis for a “‘prior’ estimate of identity.” This answer, they continue, can be justified by Bayes’s theorem, which enables us to “translate” a statistical probability into “a probability statement that describes the probative force of that statistic.”\textsuperscript{121} Somehow, in short,

\textsuperscript{118} Collins (note 000 above), 37 (“conservative estimates”), 41 (“mathematical proof”).

\textsuperscript{119} However, the prosecution was unable to get the witnesses together a second time, so this new trial never took place. Fisher 2006, pp. 21-2.

\textsuperscript{120} Collins (note 000 above), 40.

\textsuperscript{121} Finkelstein and Farley 1969-70, p. 498.
Bayes’s Theorem is to transform the statistical sense of “probable” into the epistemic. Of course, a theorem of the probability calculus, which is just a bit of mathematics, couldn’t possibly perform such a miracle of “translation.” So it comes as no great surprise that, in fact, Finkelstein and Farley bridge the gap between statistical and epistemic probabilities, not (as advertised) by applying Bayes’s Theorem, but by tinkering with the interpretation of “probable.” Like Kadane and Schum, what Finkelstein and Farley mean by “subjective” probability is the probability assigned by some subject to the proposition in question. Unlike Kadane and Schum, however, Finkelstein and Farley make an effort to explain what the latter probability is: “the relative frequency of guilt over cases judged to be similar in the degree of belief they engender.”

I think what this means is that the “subjective” probability [i.e., the objective probability assigned by some subject x] that the defendant is guilty is the proportion of [presumably, possible-but-not-necessarily-actual] cases in which the facts are different from this one but in which x judges that his degree of belief in the defendant’s guilt would be the same, in which the defendants would be guilty. Finkelstein and Farley admit their gloss is “artificial”; but the fact is, it’s close to unintelligible. What class of possible cases is a juror supposed to be imagining? How are these possible cases to be individuated? How is a juror supposed to estimate the proportion of those possible cases in which the defendant would be guilty? And how is all this supposed to work when you extrapolate it from x’s judgment of the probability of the defendant’s guilt to his assignments of prior probabilities, or of conditional probabilities?

To show that my account does better, let me begin with Justice Sullivan’s observation—with which Finkelstein and Farley agree—that statistical evidence is about a population and so, without other evidence, can’t warrant a conclusion about an

122 Id., p 504.

123 Id., p.504.
individual. Imagine a case like *Collins*, but with real, not invented, statistical evidence. It is almost, but perhaps not quite, too obvious to need saying that the statistics, by themselves, would have no bearing were it not for (i) the eyewitness evidence and (ii) the fact that the defendants fit the eyewitnesses’ description of the perpetrators. Suppose, then, that this were all the evidence we had:

- E1: eyewitnesses identify the perpetrators as a blond woman and a black man with a mustache and beard, who drove away in a yellow car;
- E2: couples fitting this description are very rare (one in $n$) in the population in the area where the robbery took place;
- E3: the Collinses fit this description.

This evidence gives *some* degree of support to the conclusion that the Collinses were the perpetrators: that the Collinses committed the crime would explain what the eyewitnesses saw, and the statistical evidence indicates that relatively few other possible explanations are possible. Supportiveness alone, however, is not enough; and the degree of warrant this evidence gives the conclusion is obviously quite low: not only is the degree of supportiveness less than overwhelming, but there are also issues about independent security (e.g., the reliability of the eyewitnesses), and the comprehensiveness of the evidence is sadly lacking—there is a good deal of obviously relevant evidence missing.

Now imagine that we have some additional evidence:

- E4: the eyewitnesses aren’t visually impaired, got a good look, have no reason to lie, etc.;
- E5: there is reason to believe the perpetrators were local;
- E6: the Collinses have no alibi;
- E7: the Collinses subsequent behavior was evasive, etc.

The addition of E4-E7 improves the explanatory integration of evidence-plus-conclusion appreciably: the conjunction of E1-E3 supports the conclusion that the Collinses committed the crime to a fairly modest degree, but the conjunction of E1-E7 supports it to
a significantly higher degree. Moreover, some worries about independent security are resolved; and E1-E7 is significantly more comprehensive than E1-E3, and the additional evidence no less favorable to the conclusion. So E1-E7 warrant the conclusion to a significantly higher degree than E1-E3 alone.

But now imagine that we have, instead, this very different additional evidence:

- E4*: the eyewitnesses are visually impaired, or didn’t get a good look, and/or have some motive to finger Janet and Malcolm Collins, and/or had seen them in handcuffs at the police station before they picked them out of the lineup, or etc.;
- E5*: there is reason to believe the perpetrators weren’t local, but from another state;
- E6*: the Collines offer an alibi, which has been confirmed;
- E7*: the Collines produce documents (verified as legitimate) showing that they bought their yellow car only after the crime took place.

E4*, E5*, E6*, and E7* undermine the conclusion that the Collines were the perpetrators; so that the conjunction of E1-E3 with E4*-E7* is better explanatorily integrated with, and hence supports, the conclusion that the Collines were not the perpetrators, rather than the conclusion that they were. And the evidence that the alibi has been confirmed and that the documents regarding the purchase of the car have been verified raises the degree of warrant of the conclusion that they were not the perpetrators by resolving some issues about the independent security of evidence undermining the conclusion that they were.

I could go on; but this is enough, I hope, to illustrate the advantages of a foundherentist approach over Finkelstein and Farley’s impressively clever, but ultimately baffling, probabilistic analysis. And if so—since Finkelstein and Farley are concerned, not just with Collins but with the role of statistical identification evidence generally—my approach should shed some light on this broader field.

When Finkelstein and Farley’s paper was published, DNA “fingerprinting” had not
yet entered the legal system.\footnote{DNA fingerprinting was first used in a criminal case in the U.S. in 1987, in the Florida case of Tommy Lee Andrews. See \textit{Andrews v. State}, 533 So.2d 841 (Fla. App. 1988).} But cases involving this increasingly common kind of evidence can illustrate the general epistemological point particularly well. Mr. Raymond Easton was arrested for a robbery on the basis of a DNA “cold hit”; statistically, the probability was very low that the match between Mr. Easton’s DNA (on file after an arrest for domestic violence) and DNA found at the crime scene was random. But Mr. Easton, who suffered from Parkinson’s disease, was too weak to dress himself or walk more than a few yards—let alone to drive to the crime scene, or to commit the crime.\footnote{\textsc{GeneWatch} UK Jan. 2005, p. 23.}

This is a case structurally much like my second imaginary extrapolation of \textit{Collins}: statistical evidence seems to support the conclusion that Mr. Easton did it (since Mr. Easton’s being at the scene would explain this DNA’s being there); but there is other evidence undermining this conclusion (since Mr. Easton was so physically impaired, we have no explanation of how he came to be at the scene, or how he could have committed the robbery). In short, the conjunction of the statistical evidence with the rest supports the conclusion that Mr. Easton \textit{didn’t} do it, and undermines the conclusion that he did. And so—assuming reasonable independent security (e.g., a reputable doctor has confirmed that Mr. Easton really has Parkinson’s), and assuming that no significant evidence is missing (e.g., that Mr. Easton paid an accomplice to drive him to the crime scene, ...., etc.), the conclusion that he didn’t do it is well warranted.

* 

Statistical evidence is not, of course, relevant only to criminal cases. DNA evidence can play a role in paternity or inheritance cases, for example; and epidemiological evidence plays a large role in toxic-tort litigation. In any case—do I really need to say this again?—my “epistemological dissent” from legal probabilism doesn’t apply only to cases
involving statistical evidence, but is quite general. The point is that, rather than enhancing our understanding of what legal degrees of proof are, probabilism impedes it—and that sound epistemology can help where legal probabilism hinders.\textsuperscript{126}

\textsuperscript{126} My thanks to Mark Migotti for his helpful comments on a draft, and to Pamela Lucken for skilled assistance in finding relevant materials.


University Press.


