

Module 9: Probability and Litigation

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.

— Judge Richard A. Posner (“An Economic Approach to the Law of Evidence,” *Stanford Law Review*, 1999)

Abstract: This module explores the connection between statements that involve probabilities and those phrases used for evidentiary purposes in the courts. We begin with Jack Weinstein, a federal judge in the Eastern District of New York, and his views on the place that probability has in litigation. Jack Weinstein may be the only federal judge ever to publish an article in a major statistics journal; his primary interests center around subjective probability and how these relate, among others, to the four levels of a “legal burden of proof”: preponderance of the evidence; clear and convincing evidence; clear, unequivocal, and convincing evidence; and proof beyond a reasonable doubt. The broad topic area of probability scales and rulers is discussed in relation to several more specific subtopics: Jeremy Bentham and his suggestion of a “persuasion thermometer”; some of Jack Weinstein’s legal rulings where probabilistic assessments were made: the cases of Vincent Gigante, Agent Orange, and Daniel Fatico. An appendix gives a redacted Weinstein opinion in this later Fatico case. Two other appendices are also given: the text of Maimonides’ 290th Negative Commandment, and a District of Columbia Court of Appeals opinion “In re As.H” (2004) that dealt with the assignment of subjective probabilities and various attendant verbal phrases in eyewitness testimony.

Contents

1	Federal Judge Jack Weinstein, Eastern District of New York (Brooklyn)	2
2	Probability Scales and Rulers	6
2.1	The Cases of Vincent Gigante and Agent Orange . . .	9
3	Appendix: Maimonides’ 290th Negative Commandment	10
4	Appendix: The Redacted Text of Judge Weinstein’s Opinion in United States v. Fatico (1978)	12
5	Appendix: District of Columbia Court of Appeals, “In re As.H” (2004)	15
1	Federal Judge Jack Weinstein, Eastern District of New York (Brooklyn)	

The retirement of Supreme Court Justice John Paul Stevens in 2010 gave President Obama a second opportunity to nominate a successor who drew the ire of the Republican Party during the confirmation process (similar to the previous such hearing with “the wise Latina woman”). For one who might have enjoyed witnessing a collective apoplexy from the conservative right, we could have suggested that President Obama nominate Jack Weinstein, a sitting federal judge in the Eastern District of New York (Brooklyn), if it weren’t for the fact that at 89 he was only one year younger than the retiring Stevens.

Weinstein is one of the most respected and influential judges in America. He has directly organized and presided over some of the most important mass tort cases of the last forty years (for example, Agent Orange, asbestos, tobacco, breast implants, DES, Zyprexa, handgun regulation, and repetitive-stress injuries).¹ For present purposes, our interest is in Weinstein's deep respect for science-based evidence in the judicial process, and in particular, for how he views probability and statistics as an intimate part of that process. He also may be the only federal judge ever to publish an article in a major statistics journal (*Statistical Science*, 1988, 3, 286–297, “Litigation and Statistics”). This last work developed out of Weinstein's association in the middle 1980s with the National Academy of Science's Panel on Statistical Assessment as Evidence in the Courts. This panel produced the comprehensive Springer-Verlag volume, *The Evolving Role of Statistical Assessments as Evidence in the Courts* (1988; Stephen E. Fienberg, Editor).

The importance that Weinstein gives to the role of probability and statistics in the judicial process is best expressed by Weinstein himself (we quote from his *Statistical Science* article):

The use of probability and statistics in the legal process is not unique to our times. Two thousand years ago, Jewish law, as stated in the Talmud, cautioned about the use of probabilistic inference. The medieval Jewish commentator Maimonides summarized this traditional view in favor of certainty when he noted:

“The 290th Commandment is a prohibition to carry out punishment on a high probability, even close to certainty . . . No punishment [should] be

¹A tort is a civil wrong; tort law concerns situations where a person's behavior has harmed someone else.

carried out except where ... the matter is established in certainty beyond any doubt ... ”

That view, requiring certainty, is not acceptable to the courts. We deal not with the truth, but with probabilities, in criminal as well as civil cases. Probabilities, express and implied, support every factual decision and inference we make in court. (p. 287)

Maimonides’ description of the 290th Negative Commandment is given in its entirety in an appendix to this module. According to this commandment, an absolute certainty of guilt is guaranteed by having two witnesses to exactly the same crime. Such a probability of guilt being identically one is what is meant by the contemporary phrase “without any shadow of a doubt.”

Two points need to be emphasized about this Mitzvah (Jewish commandment). One is the explicit unequalness of costs attached to the false positive and negative errors: “it is preferable that a thousand guilty people be set free than to execute one innocent person.” The second is in dealing with what would now be characterized as the (un)reliability of eyewitness testimony. Two eyewitnesses are required, neither is allowed to make just an inference about what happened but must have observed it directly, and exactly the same crime must be observed by both eyewitnesses. Such a high standard of eyewitness integrity might have made the current rash of DNA exonerations unnecessary.

Judge Weinstein’s interest in how probabilities could be part of a judicial process goes back some years before the National Research Council Panel. In one relevant opinion from 1978, *United States v. Fatico*, he wrestled with how subjective probabilities might be related to the four levels of a “legal burden of proof”; what level was

required in this particular case; and, finally, was it then met. The four (ordered) levels are: preponderance of the evidence; clear and convincing evidence; clear, unequivocal, and convincing evidence; and proof beyond a reasonable doubt. The case in point involved proving that Daniel Fatico was a “made” member of the Gambino organized crime family, and thus could be given a “Special Offender” status. “The consequences of (such) a ‘Special Offender’ classification are significant. In most cases, the designation delays or precludes social furloughs, release to half-way houses and transfers to other correctional institutions; in some cases, the characterization may bar early parole” (text taken from the opinion). The summary of Weinstein’s final opinion in the Fatico case follows:

In view of prior proceedings, the key question of law now presented is what burden of proof must the government meet in establishing a critical fact not proved at a criminal trial that may substantially enhance the sentence to be imposed upon a defendant. There are no precedents directly on point.

The critical factual issue is whether the defendant was a “made” member of an organized crime family. Clear, unequivocal and convincing evidence adduced by the government at the sentencing hearing establishes this proposition of fact.

The text of Weinstein’s opinion in the Fatico case explains some of the connections between subjective probabilities, burdens of proof, and the need for different levels depending on the particular case (we note that the numerical values suggested in this opinion as corresponding to the various levels of proof, appear to be based only on Judge Weinstein’s “best guesses”). We redact part of his opinion in an appendix to this module.

Other common standards used for police searches or arrests might

also be related to an explicit probability scale. The lowest standard (perhaps a probability of 20%) would be “reasonable suspicion” to determine whether a brief investigative stop or search by any governmental agent is warranted (in the 2010 “Papers, Please” law in Arizona, a “reasonable suspicion” standard is set for requesting documentation). A higher standard would be “probable cause” to assess whether a search or arrest is warranted, or whether a grand jury should issue an indictment. A value of, say, 40% might indicate a “probable cause” level that would put it somewhat below a “preponderance of the evidence” criterion. In all cases, a mapping of such verbal statements to numerical values requires “wiggle” room for vagueness, possibly in the form of an interval estimate rather than a point estimate. As an example of this variability of assessment in the Fatico case, Judge Weinstein informally surveyed the judges in his district court regarding the four different standards of proof. The data are given in Table 1 (taken from Fienberg, 1988, p. 204). We leave it to you to decide whether you would want Judge 4 or 7 to hear your case.

2 Probability Scales and Rulers

The topic of relating a legal understanding of burdens of proof to numerical probability values has been around for a very long time. Fienberg (1988, p. 212) provides a short discussion of Jeremy Bentham’s (1827) suggestion of a “persuasion thermometer,” and some contemporary reaction to this idea from Thomas Starkie (1833).² We

²This is the same Bentham known for utilitarianism, and more amusingly, for the “auto-
icon.” A short section from the Wikipedia article on “Jeremy Bentham” describes the

Table 1: Probabilities associated with different standards of proof by judges in the Eastern District of New York.

Judge	Preponderance	Clear and convincing	Clear, unequivocal, and convincing	Beyond a reasonable doubt	row median
1	50+	60	70	85	65
2	51	65	67	90	66
3	50+	60-70	65-75	80	67
4	50+	67	70	76	69
5	50+	Standard is elusive		90	70
6	50+	70+	70+	85	70
7	50+	60	90	85	72
8	50+	70+	80+	95+	75
9	50.1	75	75	85	75
10	51	Cannot estimate			–
column median	50	66	70	85	

quote:

Jeremy Bentham appears to have been the first jurist to seriously propose auto-icon:

As requested in his will, Bentham’s body was dissected as part of a public anatomy lecture. Afterward, the skeleton and head were preserved and stored in a wooden cabinet called the “Auto-icon,” with the skeleton stuffed out with hay and dressed in Bentham’s clothes. Originally kept by his disciple Thomas Southwood Smith, it was acquired by University College London in 1850. It is normally kept on public display at the end of the South Cloisters in the main building of the college, but for the 100th and 150th anniversaries of the college, it was brought to the meeting of the College Council, where it was listed as “present but not voting.”

The Auto-icon has a wax head, as Bentham’s head was badly damaged in the preservation process. The real head was displayed in the same case for many years, but became the target of repeated student pranks, including being stolen on more than one occasion. It is now locked away securely.

that witnesses and judges numerically estimate their degrees of persuasion. Bentham (1827; Vol. 1, pp. 71–109) envisioned a kind of moral thermometer:

The scale being understood to be composed of ten degrees—in the language applied by the French philosophers to thermometers, a decigrade scale—a man says, My persuasion is at 10 or 9, etc. affirmative, or at least 10, etc. negative . . .

Bentham’s proposal was greeted with something just short of ridicule, in part on the pragmatic grounds of its inherent ambiguity and potential misuse, and in part on the more fundamental ground that legal probabilities are incapable of numerical expression. Thomas Starkie (1833) was merely the most forceful when he wrote:

The notions of those who have supposed that mere moral probabilities or relations could ever be represented by numbers or space, and thus be subjected to arithmetical analysis, cannot but be regarded as visionary and chimerical. (p. 212)

Several particularly knotty problems and (mis)interpretations when it comes to assigning numbers to the possibility of guilt arise most markedly in eyewitness identification. Because cases involving eyewitness testimony are typically criminal cases, they demand burdens of proof “beyond a reasonable doubt”; thus, the (un)reliability of eyewitness identification becomes problematic when it is the primary (or only) evidence presented to meet this standard. As discussed extensively in the judgment and decision-making literature, there is a distinction between making a subjective estimate of some quantity, and one’s confidence in that estimate once made. For example, suppose someone picks a suspect out of a lineup, and is then asked the (Bentham) question, “on a scale of from one to ten, characterize your level of ‘certainty’.” Does an answer of “seven or eight” translate into a probability of innocence of two or three out of ten? Exactly such confusing situations, however, arise. We give a fairly

extensive redaction in an appendix to this module of an opinion from the District of Columbia Court of Appeals in a case named “In re As.H” (2004). It combines extremely well both the issues of eyewitness (un)reliability and the attempt to quantify that which may be better left in words; the dissenting Associate Judge Farrel noted pointedly: “I believe that the entire effort to quantify the standard of proof beyond a reasonable doubt is a search for fool’s gold.”

2.1 The Cases of Vincent Gigante and Agent Orange

Although Judge Weinstein’s reputation may rest on his involvement with mass toxic torts, his most entertaining case occurred in the middle 1990s, with the murder-conspiracy and racketeering trial and conviction of Vincent Gigante, the boss of the most powerful Mafia family in the United States. The issue here was assessing the evidence of Gigante’s mental fitness to be sentenced to prison, and separating such evidence from the putative malingering of Gigante. Again, Judge Weinstein needed to evaluate the evidence and make a probabilistic assessment (“beyond a reasonable doubt”) that Gigante’s trial was a “valid” one.

Apart from the great legal theater that the Gigante case provided, Judge Weinstein’s most famous trials all involve the Agent Orange defoliant used extensively by the United States in Vietnam in the 1960s. Originally, he oversaw in the middle 1980s the \$200 million settlement fund provided by those companies manufacturing the agent. Most recently, Judge Weinstein presided over the dismissal of the civil lawsuit filed on behalf of millions of Vietnamese individuals. The 233-page decision in this case is an incredible “read” about

United States polices during this unfortunate period in our country's history. The suit was dismissed not because of poor documentation of the effects of Agent Orange and various ensuing conditions, but because of other legal conditions. The judge concluded that even if the United States had been a Geneva Accord signatory (outlawing use of poisonous gases during war), Agent Orange would not have been banned: "The prohibition extended only to gases deployed for their asphyxiating or toxic effects on man, not to herbicides designed to affect plants that may have unintended harmful side effects on people" (In re "Agent Orange" Product Liability Litigation, 2005, p. 190). The National Academy of Science through its Institute of Medicine, regularly updates what we know about the effects of Agent Orange, and continues to document many associations between it and various disease conditions. The issues in the Vietnamese lawsuit, however, did not hinge on using a probability of causation assessment, but rather on whether, given the circumstances of the war, the United States could be held responsible for what it did in Vietnam in the 1960s.

3 Appendix: Maimonides' 290th Negative Commandment

"And an innocent and righteous person you shall not slay" — Exodus 23:7.

Negative Commandment 290

Issuing a Punitive Sentence Based on Circumstantial Evidence:

The 290th prohibition is that we are forbidden from punishing someone based on our estimation [without actual testimony], even if his guilt is virtually certain. An example of this is a person who was chasing after his enemy to kill him. The pursued escaped into a house and the pursuer entered the

house after him. We enter the house after them and find the victim lying murdered, with the pursuer standing over him holding a knife, with both covered with blood. The Sanhedrin may not inflict the death penalty on this pursuer since there were no witnesses who actually saw the murder.

The Torah of Truth (Toras Emess) comes to prohibit his execution with G—d’s statement (exalted be He), “Do not kill a person who has not been proven guilty.”

Our Sages said in *Mechilta*: “If they saw him chasing after another to kill him and they warned him, saying, ‘He is a Jew, a son of the Covenant! If you kill him you will be executed!’ If the two went out of sight and they found one murdered, with the sword in the murderer’s hand dripping blood, one might think that he can be executed. The Torah therefore says, ‘Do not kill a person who has not been proven guilty.’”

Do not question this law and think that it is unjust, for there are some possibilities that are extremely probable, others that are extremely unlikely, and others in between. The category of “possible” is very broad, and if the Torah allowed the High Court to punish when the offense was very probable and almost definite (similar to the above example), then they would carry out punishment in cases which were less and less probable, until people would be constantly executed based on flimsy estimation and the judges’ imagination. G—d (exalted be He), therefore “closed the door” to this possibility and forbid any punishment unless there are witnesses who are certain beyond a doubt that the event transpired and that there is no other possible explanation.

If we do not inflict punishment even when the offense is most probable, the worst that could happen is that someone who is really guilty will be found innocent. But if punishment was given based on estimation and circumstantial evidence, it is possible that someday an innocent person would be executed. And it is preferable and more proper that even a thousand guilty people be set free than to someday execute even one innocent person.

Similarly, if two witnesses testified that the person committed two capital offenses, but each one saw only one act and not the other, he cannot be executed. For example: One witness testified that he saw a person doing a *melachah* on *Shabbos* and warned him not to. Another witness testified that

he saw the person worshipping idols and warned him not to. This person cannot be executed by stoning. Our Sages said, “If one witness testified that he worshipped the sun and the other testified that he worshipped the moon, one might think that they can be joined together. The Torah therefore said, ‘Do not kill a person who has not been proven guilty.’”

4 Appendix: The Redacted Text of Judge Weinstein’s Opinion in *United States v. Fatico* (1978)

We begin with the caution of Justice Brennan in *Speiser v. Randall*, about the crucial nature of fact finding procedures:

To experienced lawyers it is commonplace that the outcome of a lawsuit and hence the vindication of legal rights depends more often on how the factfinder appraises the facts than on a disputed construction of a statute or interpretation of a line of precedents. Thus the procedures by which the facts of the case are determined assume an importance fully as great as the validity of the substantive rule of law to be applied. And the more important the rights at stake, the more important must be the procedural safeguards surrounding those rights.

The “question of what degree of proof is required . . . is the kind of question which has traditionally been left to the judiciary to resolve . . . ”

Broadly stated, the standard of proof reflects the risk of winning or losing a given adversary proceeding or, stated differently, the certainty with which the party bearing the burden of proof must convince the factfinder.

As Justice Harlan explained in his concurrence in *Winship*, the choice of an appropriate burden of proof depends in large measure on society’s assessment of the stakes involved in a judicial proceeding.

In a judicial proceeding in which there is a dispute about the facts of some earlier event, the factfinder cannot acquire unassailably accurate knowledge of what happened. Instead, all the factfinder can acquire is a belief of what Probably happened. The intensity of this belief—the degree to which a factfinder is convinced that a given act actually occurred—can, of course, vary. In this regard, a standard of proof represents an attempt 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. Although the phrases “preponderance of the evidence” and “proof beyond a reasonable doubt” are quantitatively imprecise, they do communicate to the finder of fact different notions concerning the degree of confidence he is expected to have in the correctness of his factual conclusions.

Thus, the burden of proof in any particular class of cases lies along a continuum from low probability to very high probability.

Preponderance of the Evidence:

As a general rule, a “preponderance of the evidence” [or] more probable than not standard, is relied upon in civil suits where the law is indifferent as between plaintiffs and defendants, but seeks to minimize the probability of error.

In a civil suit between two private parties for money damages, for example, we view it as no more serious in general for there to be an erroneous verdict in the defendant’s favor than for there to be an erroneous verdict in the plaintiff’s favor. A preponderance of the evidence standard therefore seems peculiarly appropriate; as explained most sensibly, it simply requires the trier of fact “to believe that the existence of a fact is more probable than its nonexistence before (he) may find in favor of the party who has the burden to persuade the (judge) of the fact’s existence.”

Quantified, the preponderance standard would be 50+% Probable.

Clear and Convincing Evidence:

In some civil proceedings where moral turpitude is implied, the courts utilize the standard of “clear and convincing evidence,” a test somewhat stricter than preponderance of the evidence.

Where proof of another crime is being used as relevant evidence pursuant to Rules 401 to 404 of the Federal Rules of Evidence, the most common test articulated is some form of the “clear and convincing” standard.

Quantified, the probabilities might be in the order of above 70% under a clear and convincing evidence burden.

Clear, Unequivocal and Convincing Evidence:

“In situations where the various interests of society are pitted against restrictions on the liberty of the individual, a more demanding standard is frequently imposed, such as proof by clear, unequivocal and convincing evi-

dence.” The Supreme Court has applied this stricter standard to deportation proceedings, denaturalization cases, and expatriation cases. In *Woodby*, the Court explained:

To be sure, a deportation proceeding is not a criminal prosecution. But it does not syllogistically follow that a person may be banished from this country upon no higher degree of proof than applies in a negligence case. This Court has not closed its eyes to the drastic deprivations that may follow when a resident of this country is compelled by our Government to forsake all the bonds formed here and go to a foreign land where he often has no contemporary identification.

In terms of percentages, the probabilities for clear, unequivocal and convincing evidence might be in the order of above 80% under this standard.

Proof Beyond a Reasonable Doubt:

The standard of “proof beyond a reasonable doubt” is constitutionally mandated for elements of a criminal offense. Writing for the majority in *Winship*, Justice Brennan enumerated the “cogent reasons” why the “‘reasonable-doubt’ standard plays a vital role in the American scheme of criminal procedure” and “is a prime instrument for reducing the risk of convictions resting on factual error.”

The accused during a criminal prosecution has at stake interest of immense importance, both because of the possibility that he may lose his liberty upon conviction and because of the certainty that he would be stigmatized by the conviction. Accordingly, a society that values the good name and freedom of every individual should not condemn a man for commission of a crime when there is reasonable doubt about his guilt. As we said in *Speiser v. Randall*, “There is always in litigation a margin of error, representing error in fact finding, which both parties must take into account. Where one party has at stake an interest of transcending value as a criminal defendant—his liberty—this margin of error is reduced as to him by the process of placing on the other party the burden of . . . persuading the factfinder at the conclusion of the trial of his guilt beyond a reasonable doubt. Due process commands that no man shall lose his liberty unless the Government has borne the burden of . . . convincing the factfinder of his guilt.” . . .

Moreover, use of the reasonable-doubt standard is indispensable to com-

mand the respect and confidence of the community in applications of the criminal law. It is critical that the moral force of the criminal law not be diluted by a standard of proof that leaves people in doubt whether innocent men are being condemned.

In capital cases, the beyond a reasonable doubt standard has been utilized for findings of fact necessary to impose the death penalty after a finding of guilt.

Many state courts, in interpreting state recidivism statutes, have held that proof of past crimes must be established beyond a reasonable doubt.

In civil commitment cases, where the stakes most resemble those at risk in a criminal trial, some courts have held that the beyond a reasonable doubt standard is required.

If quantified, the beyond a reasonable doubt standard might be in the range of 95+% Probable.

5 Appendix: District of Columbia Court of Appeals, “In re As.H” (2004)

DISTRICT OF COLUMBIA COURT OF APPEALS IN RE AS.H.

SCHWELB, Associate Judge: This juvenile delinquency case is more than five years old. On January 20, 1999, following a factfinding hearing, As.H., then sixteen years of age, was adjudicated guilty of robbery. The sole evidence implicating As.H. in the offense was the testimony of the victim, Ms. Michal Freedhoff, who identified As.H. at a photo array almost a month after the robbery and again in court more than four months after that. Ms. Freedhoff described her level of certainty on both occasions, however, as “seven or eight” on a scale of one to ten. Because Ms. Freedhoff was obviously less than positive regarding her identification, and for other reasons described below, we conclude as a matter of law that the evidence was insufficient to prove beyond a reasonable doubt that As.H. was involved in the robbery. Accordingly, we reverse.

I. In the early morning hours of August 17, 1998, between 12:30 and 1:00 a.m., Ms. Freedhoff was robbed by three or more young men. The assailants

knocked Ms. Freedhoff to the ground, threatened her with “a long piece of wood” which, Ms. Freedhoff believed, was “suppose[d] to look like a rifle,” ordered her to “shut up, bitch,” and robbed her of her purse and her personal electronic organizer. Ms. Freedhoff promptly reported the crime to the police. Officers detained a group of young men shortly after the robbery and arranged a show-up, but Ms. Freedhoff stated that the detained individuals were not the robbers. Indeed, she was “completely” certain that the individuals at the show-up were not the guilty parties.

Ms. Freedhoff testified that there were street lights in the area where the robbery occurred. She further stated that she had been outside in the street for some time, so that her eyes had become accustomed to the dark. Nevertheless, Ms. Freedhoff could not provide an informative description of her assailants. According to Detective Ross, she recalled nothing distinctive about their clothing; “young black males and baggy clothes” was his recollection of her report. At the factfinding hearing, which took place more than five months after the robbery, Ms. Freedhoff recalled that the robbers were teenagers, “two dark-skinned and one light,” each of a different height, and that “one had shorts and sneakers and another may have had a hat.” Ms. Freedhoff was also uncertain as to the role which the individual she tentatively identified as As.H. allegedly played in the robbery.

On September 11, 1998, Detective Ross showed Ms. Freedhoff an array of nine polaroid pictures and asked her if she recognized anyone who was involved in the offense. At a hearing on As.H.’s motion to suppress identification, Ms. Freedhoff testified as follows regarding this array:

Q: Now, Ms. Freedhoff, on that day did you identify any of the people in the photos as having been involved in the incident of August 16th?

A: Yes, I did.

Q: Which photos did you identify?

A: These two marked nine and [ten] I was very certain about and the two marked three and four I was less certain about.

Q: During the identification procedure, did you talk to the detective about your level of certainty?

A: Yes.

Q: In terms of nine and [ten], what was your level of certainty that those

people were involved?

A: I [was] asked to rate them on a scale of—I believe it was one to [ten]—and I believe I said it was, that nine and [ten], I was seven or eight.

Q: And in terms of three and four, how did you rate those?

A: Six.

According to Ms. Freedhoff, the photograph of As.H. was No. 10. At the factfinding hearing, Ms. Freedhoff initially stated that she saw one of the robbers sitting in the courtroom, pointing out As.H. When asked which of the individuals in the array he was, Ms. Freedhoff “believed” that it “would be Number 10.” However, when counsel for the District of Columbia again asked Ms. Freedhoff about her present level of certainty in making the identification—how certain are you?—the witness adhered to her previous estimate: “At the time, on a scale of one to [ten], I said that I was seven or eight.”

According to Detective Ross, who also testified regarding the viewing of the photo array, Ms. Freedhoff was “comfortable in saying they could be the people that robbed her.” Ross further disclosed that he “may have discussed with [Ms. Freedhoff] that I had a previous history with the persons that she had picked. They were my possible suspects in the case.”

Without elaborating on his reasons, the trial judge denied As.H.’s motion to suppress identification and found As.H. guilty as charged. This appeal followed.

II. In evaluating claims of evidentiary insufficiency in juvenile delinquency appeals, we view the record “in the light most favorable to the [District], giving full play to the right of the judge, as the trier of fact, to determine credibility, weigh the evidence, and draw reasonable inferences . . . We will reverse on insufficiency grounds only when the [District] has failed to produce evidence upon which a reasonable mind might fairly find guilt beyond a reasonable doubt.” “Even identification testimony of a single eyewitness will be sufficient so long as a reasonable person could find the identification convincing beyond a reasonable doubt.” Moreover, the District was not required to prove As.H.’s guilt beyond all doubt. “There is no rule of law which requires an identification to be positive beyond any shadow of doubt.”

Nevertheless, the “[beyond a] reasonable doubt” standard of proof is a

formidable one. It “requires the factfinder to reach a subjective state of near certitude of the guilt of the accused.” Although appellate review is deferential, we have “the obligation to take seriously the requirement that the evidence in a criminal prosecution must be strong enough that a jury behaving rationally really could find it persuasive beyond a reasonable doubt.” Moreover, “while [the trier of fact] is entitled to draw a vast range of reasonable inferences from evidence, [he or she] may not base [an adjudication of guilt] on mere speculation.”³

In the present case, we have an eyewitness identification of questionable certitude, and the witness and the respondent are strangers. Ms. Freedhoff saw her assailants at night and under extremely stressful conditions. Moreover, this is a “pure” eyewitness identification case; there is no evidence linking As.H. to the robbery except for Ms. Freedhoff’s statements upon viewing the array almost a month after the event and her testimony at the factfinding hearing more than five months after she was robbed.

The vagaries of eyewitness identification, and the potential for wrongful convictions or adjudications based upon such evidence, have long been recognized in the District of Columbia. More recently, in *Webster v. United States*, we summarized this concern as follows:

“[T]he identification of strangers is proverbially untrustworthy. The hazards of such testimony are established by a formidable number of instances in the records of English and American trials.” FELIX FRANKFURTER, *THE CASE OF SACCO AND VANZETTI* (1927). Indeed, “[p]ositive identification of a person not previously known to the witness is perhaps the most fearful testimony known to the law of evidence.” Even if the witness professes certainty, “it is well recognized that the most positive eyewitness is not necessarily the most reliable.”

Here, the witness did not even profess certainty. Moreover, the present case concerns a hesitant cross-racial identification by a white woman of a

³The court emphasized in *Crawley* that as appellate judges, we have the responsibility in eyewitness identification cases “to draw upon our own experience, value judgments, and common sense in determining whether the [finding] reached was in keeping with the facts.” Although this observation might be viewed today as an unduly activist formulation of an appellate court’s function, it illustrates the concern of conscientious judges regarding the possibility that a mistaken identification may send an innocent person to prison.

black teenager, and “[i]t is well established that there exists a comparative difficulty in recognizing individual members of a race different from one’s own.” ELIZABETH LOFTUS, *EYEWITNESS TESTIMONY*; see *State v. Cromedy*, (discussing at length the difficulties in cross-racial identification and mandating a jury instruction on the subject in some cases); John P. Rutledge, *They All Look Alike: The Inaccuracy of Cross-Racial Identifications*, 28 AM. J. CRIM. L. 207 (2001).

It is in the context of these realities that we now turn to the dispositive issue in this appeal, namely, whether Ms. Freedhoff’s testimony—the only evidence of As.H.’s participation in the robbery—was legally sufficient to support a finding of guilt beyond a reasonable doubt. The key fact is that, both when viewing the polaroid photographs and when testifying in open court, Ms. Freedhoff candidly characterized her level of “certainty”—i.e., of her being “very certain”—as seven or eight on a scale of one to ten. Her testimony leads inexorably to the conclusion that her level of uncertainty—i.e., the possibility that As.H. was not involved—was two or three out of ten—a 20% to 30% possibility of innocence. This differs dramatically from Ms. Freedhoff’s complete certainty that the young men she viewed at the show-up on the night of the offense were not the robbers. The contrast between Ms. Freedhoff’s statements in the two situations is revealing, and surely negates the “near certitude” that is required for a showing of guilt beyond a reasonable doubt. The “seven or eight out of ten” assessment is also consistent with Detective Ross’ recollection of Ms. Freedhoff’s account: As.H. and others “could be the people that robbed her,” and As.H. “looked like” one of the kids. It is, of course, difficult (if not impossible) to place a meaningful numerical value on reasonable doubt. See generally Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV 1329 (1971); Underwood, *The Thumb on the Scales of Justice; Burden of Persuasion in Criminal Cases*, 86 Yale L.J. 1299, 1309–11 (1977) (hereinafter Underwood). Professor Wigmore cites a study in which judges in Chicago were asked to:

translate into probability statements their sense of what it means to be convinced by a preponderance of the evidence, and to be convinced beyond a reasonable doubt. When responding to questionnaires, at least, the judges

thought there was an important difference: almost a third of the responding judges put “beyond a reasonable doubt” at 100%, another third put it at 90% or 95%, and most of the rest put it at 80% or 85%. For the preponderance standard, by contrast, over half put it at 55%, and most of the rest put it between 60% and 75%.

Although the Chicago study alone is not dispositive of this appeal, it reveals that very few judges, if any, would have regarded an 80% probability as sufficient to prove guilt beyond a reasonable doubt, and that all of them would have considered a 70% probability as altogether inadequate. For the Chicago judges, Ms. Freedhoff’s “certainty” appears to be well outside the ballpark for proof in a criminal case. In *Fatico*, nine judges of the United States District Court for the Eastern District of New York, responding to a poll by Judge Weinstein, the co-author of a leading text on evidence, suggested percentages of 76%, 80%, 85%, 85%, 85%, 85%, 90%, 90% and 95%, as reflecting the standard for proof beyond a reasonable doubt. Thus, at most, two of the nine judges polled by Judge Weinstein would have found the level of assurance voiced by Ms. Freedhoff sufficient to support a finding of guilt.⁴

But, argues the District, Ms. Freedhoff “was not asked for a level of accuracy or how sure she was, but, given the certainty of her identification, how high a level of certainty she had felt.” Therefore, the argument goes, “the trier of fact can be confident that the witness felt that her identification was very certain.” We do not find this contention at all persuasive. Taken to its logical conclusion, it would mean that if Ms. Freedhoff had expressed a level of certainty of one in ten—10%—this would be sufficient to support a finding of guilt. The notion that Ms. Freedhoff was assessing varying gradations of certainty, all of them very certain, is also at odds with what she told Detec-

⁴Commenting on the same Chicago study in one of its submissions, the District reveals only that “about one-third of the judges put it at 80%-85%.” Unfortunately, by failing to mention that one third of the judges put “beyond a reasonable doubt” at 100% and that another third put it at 90%-95%, the District presents us with a misleading picture of the results of the study. Remarkably, the District then argues that we should affirm because judges who try to quantify reasonable doubt place it “not that far distant from Ms. Freedhoff’s estimate.” In fact, the contrast between the judges’ estimates and Ms. Freedhoff’s articulation is quite remarkable, and a contention that fails to take this contrast into account is necessarily fallacious.

tive Ross, namely, that As.H. “looks like” or “could have been” one of the robbers.⁵

Professor Lawrence Tribe has written:

[I]t may well be . . . that there is something intrinsically immoral about condemning a man as a criminal while telling oneself, “I believe that there is a chance of one in twenty that this defendant is innocent, but a 1/20 risk of sacrificing him erroneously is one I am willing to run in the interest of the public’s—and my own—safety.”

It may be that Professor Tribe’s proposition is more suited to the world of academe than to the less rarefied realities of the Superior Court’s criminal docket—realities in which “beyond all doubt” presents an idealist’s impossible dream, while “beyond a reasonable doubt” provides a workable standard. This case, however, is not like the hypothetical one that disturbed Professor Tribe. Here, the doubt of the sole identifying witness in a night-time robbery by strangers to her stood at two or three out of ten, or 20%-30%. We conclude, at least on this record, that this level of uncertainty constituted reasonable doubt as a matter of law. Accordingly, we reverse the adjudication of guilt and remand the case to the Superior Court with directions to enter a judgment of not guilty and to dismiss the petition.⁶

⁵The District also argues that, in open court, Ms. Freedhoff “unhesitatingly and positively” identified the respondent. As we have explained in Part I of this opinion, however, the full context of Ms. Freedhoff’s courtroom testimony reveals that, five months after the robbery, she was no more certain of her identification than she had been when she viewed the photo array. Moreover, after Ms. Freedhoff had selected photographs at the array, Detective Ross revealed that he “had a previous history with the persons she had picked,” and that they were his “possible suspects in the case.” “[W]here . . . the police consider an individual to be a possible perpetrator and a witness makes an initially ambiguous identification, there may develop a process of mutual bolstering which converts initial tentativeness into ultimate certainty.” “The victim relies on the expertise of the officer and the officer upon the victim’s identification.”

⁶Our dissenting colleague argues that reasonable doubt is not susceptible of ready quantification, and we agree. But where, as in this case, the sole identifying witness described her own level of “certainty” as only seven or eight on a scale of ten, then, notwithstanding the difficulty of quantification in the abstract, this level of unsureness necessarily raises a reasonable doubt and negates the requisite finding of “near certitude” that As.H. was one of the robbers. Nothing in this opinion holds or even remotely suggests that a cross-racial identification is insufficient as a matter of law or that the trier of fact is required to discount

So ordered.

FARRELL, Associate Judge, dissenting: Less than a month after she was assaulted by three young men, the complainant, Ms. Freedhoff, identified two men from photographs as among the assailants. One was appellant. According to the detective who showed her the photographs, she did not hesitate in picking appellant, and at the hearing on appellant's motion to suppress the identification she twice stated that she had been "very certain" in selecting his photograph. At trial, although she could not remember appellant's exact role in the assault, she stated that she had been able to see all three assailants well, that the two people she was "certain of" in her identification "were probably the two" who had been "in front of [her]" during the assault, and that she had identified them because "they looked very familiar to [her]" as being the people that were involved." Ms. Freedhoff was not given to quick accusations: at a show-up confrontation shortly after the assault, she had been "[completely] certain" that the individuals shown to her were not the assailants. The trial judge, sitting as trier of fact, found her testimony convincing and found appellant guilty beyond a reasonable doubt.

The majority sets that verdict aside. Although concededly unable to replicate Judge Mitchell's vantage point in assessing the complainant's demeanor and the strength of her belief as she recalled the robbery and identification, it concludes that the identification was too weak as a matter of law to support conviction. And it does so at bottom for one reason: when asked by the detective her level of certainty "on a scale of one to ten" in identifying appellant, Ms. Freedhoff had answered "seven or eight." This, in the majority's view, explains what she meant when she said she was "very certain," and a level of uncertainty of an uncorroborated eyewitness "st[anding] at two or three out of ten, or 20%-30%[,] . . . constituted reasonable doubt as a matter of law."

The majority thus decides that the trier of fact could not convict based on testimony of a victim who was as much as four-fifths certain of her identification. The reasonable doubt in this case arises from the witness' very limited certainty (seven or eight on a scale of ten) regarding her uncorroborated identification. The difficulties of eyewitness identification of strangers in general, as well as of cross-racial identification, provide the context in which the witness' uncertainty arose.

tification. I do not agree, basically because I believe that the entire effort to quantify the standard of proof beyond a reasonable doubt is a search for fool's gold. Ms. Freedhoff stated that she was very certain of her identification; she was questioned extensively about the circumstances of the photo display and the assault; and she offered reasons for her certainty. The fact that when asked to rate her certainty "on a scale of one to ten" she answered "seven or eight" cannot be decisive unless, like the majority, one is ready to substitute an unreliable, quantitative test of certainty for the intensely qualitative standard of proof beyond a reasonable doubt. Even in popular usage, the "scale of one to ten" as an indicator of belief is notoriously imprecise. People who in any ultimate—and unascertainable—sense probably share the same level of conviction may translate that very differently into numbers, and even the same person will change his mind from one moment to the next in assigning a percentage to his belief. Treating "one to ten" as a decisive indicator of the sufficiency of identification evidence thus elevates to a legal standard a popular measure that makes no claim at all to precision. As Wigmore stated long ago in this context, "The truth is that no one has yet invented or discovered a mode of measurement for the intensity of human belief. Hence there can be yet no successful method of communicating intelligibly . . . a sound method of self-analysis for one's belief." Here, for example, Ms. Freedhoff equated "seven or eight" with being "very certain"; for all we know, she thought that any higher number would approach mathematical or absolute certainty, something the reasonable doubt standard does not require. The trial judge wisely did not view her attempt to furnish a numerical equivalent for her belief as conclusive, and neither should we.

The judicial straw polls cited by the majority merely confirm the futility of defining a percentual range (or "ball-park," to quote the majority) within which proof beyond a reasonable doubt must lie. Had Ms. Freedhoff added five percent to her belief-assessment (as much as "85%" rather than as much as "80%"), she would have come well within the range of, for example, Judge Weinstein's survey in *Fatico*. A factfinder's evaluation of credibility and intensity of belief should not be overridden by such inexact and even trivial differences of quantification.

Another aspect of the majority's opinion requires comment. It points to

“[t]he vagaries of eyewitness identification,” explains that this was a case of “cross-racial identification by a white woman of a black teenager,” and cites to the “well established . . . comparative difficulty in recognizing individual members of a race different from one’s own.” [quoting ELIZABETH LOFTUS, EYEWITNESS TESTIMONY]. It is not clear what the majority means by this discussion. The present appeal is not about whether a trier of fact may hear expert testimony or be instructed regarding the uncertainties of eyewitness identification, cross-racial or any other. Here the majority holds the identification insufficient as a matter of law, which implies that the trier of fact was required to discount the identification to an (undefined) extent because of the intrinsic weakness of eyewitness identifications generally or because this one was cross-racial. Either basis would be unprecedented. If, as I prefer to believe, that is not what the majority intends, then I respectfully suggest that the entire discussion of the point is dictum.

I would affirm the judgment of the trial court

References

- [1] Bentham, J. (1827). *Rationale of judicial evidence, specially applied to English practice* (J. S. Mill, Ed.). London: Hunt and Clarke.
- [2] Fienberg, S. E. (Ed.). (1988). *The evolving role of statistical assessments as evidence in the courts*. New York: Springer-Verlag.
- [3] Starkie, T. (1833). *A practical treatise of the law of evidence and digest of proofs, in civil and criminal proceedings* (2nd ed.). London: J. and W. T. Clarke.