ARTICLES

DEATH OF PARADOX: THE KILLER LOGIC BENEATH THE STANDARDS OF PROOF

Kevin M. Clermont*

The prevailing but contested view of proof standards is that fact-finders should determine facts by probabilistic reasoning. Given imperfect evidence, they first should ask themselves what they think the chances are that the burdened party would be right were the truth to become known, and they then should compare those chances to the applicable standard of proof.

I contend that for understanding the standards of proof, the modern versions of logic—in particular, fuzzy logic and belief functions—work better than classical logic and probability theory. This modern logic suggests that fact-finders first assess evidence of an imprecisely perceived and described reality to form a fuzzy degree of belief in a fact's existence, and they then apply the standard of proof by comparing their belief in a fact's existence to their belief in its negation.

This understanding nicely explains how the standard of proof actually works in the law world. While conforming more closely to what we know of people's cognition, the new understanding captures better how the law formulates and manipulates the standards and it also gives a superior mental image of the fact-finders' task. One virtue of this conceptualization is that it is not a radical reconception. Another virtue is that it nevertheless manages to resolve some stubborn problems of proof, including the infamous conjunction paradox.

TABLE OF CONTENTS

*Le seul véritable voyage, le seul bain de Jouvence, ce ne serait pas d'aller vers de nouveaux paysages, mais d'avoir d'autres yeux, de voir l'univers avec les yeux d'un autre, de cent autres . . . .***

  ** 12 MARCEL PROUST, À LA RECHERCHE DU TEMPS PERDU 69 (1923). *But cf.* PIERRE BAYARD, COMMENT PARLER DES LIEUX OÙ L'ON N'A PAS ÉTÉ? (2012) ("*contrairement aux idées reçues, il est tout à fait possible d'avoir un échange passionnant à propos d'un endroit où l'on n'a jamais mis les pieds*"); PIERRE BAYARD, HOW TO TALK ABOUT BOOKS YOU HAVEN'T READ, at xvii (2007) ("In my experience . . . it's totally possible to carry on an engaging conversation about a book you haven't read . . . ."); Deborah Solomon, *My Reader, My Double*, N.Y. TIMES MAG., http://www.nytimes.com/2007/10/28/magazine/28wwln-Q4-t.html (Oct. 28, 2007) (interviewing Pierre Bayard, a Proust expert, who admits to having only skimmed Proust); *The Books We Lie About*, THE 6TH FLOOR (May 10, 2012, 5:54 PM), http://6thfloor.blogs.nytimes.com/2012/05/10/the-books-we-lie-about/ (listing *How to Talk About Books You Haven't Read* among "books we've been less than truthful about having read"); *Top Ten Books People Lie*

INTRODUCTION

We have made tremendous strides, albeit only recently, toward understanding the process of proof. The wonderful "new evidence" scholarship has made especial progress by shifting the focus of evidence scholarship from rules of admissibility to the nature of proof, while opening the door to interdisciplinary insights, including those from psychology.[1] Yet the new work has tended to remain either too wedded or overly hostile to subjective probabilities for evaluating evidence[2] and to Bayes' theorem for combining evidence,[3] and so caused the debates to become "unproductive and sterile."[4] In any event, the

---

About Reading, SUPERGIRL SAVES THE WORLD (Mar. 13, 2010), http://realsupergirl. wordpress.com/2010/03/13/top-ten-books-people-lie-about-reading (listing *In Remembrance of Things Past/À la recherche du temps perdu*).

1   *See generally* WILLIAM TWINING, RETHINKING EVIDENCE 237–48 (2d ed. 2006) ("[I]t is illuminating to view questions about evidence and proof as questions about the processing and uses of information in important decisions in litigation."); Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof, in* PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE 61, 61 (Peter Tillers & Eric D. Green eds., 1988) ("Evidence is being transformed from a field concerned with the articulation of rules to a field concerned with the process of proof. . . . [D]isciplines outside the law, like mathematics, psychology and philosophy, are being plumbed for the guidance they can give."); Roger C. Park & Michael J. Saks, *Evidence Scholarship Reconsidered: Results of the Interdisciplinary Turn,* 47 B.C. L. REV. 949, 949 (2006) ("[T]he changing field of evidence scholarship . . . has become decidedly interdisciplinary."). Although "new," this work represented the necessary return to abandoned efforts by past greats such as Wigmore. *See, e.g.,* TERENCE ANDERSON ET AL., ANALYSIS OF EVIDENCE (2d ed. 2005) (building on JOHN HENRY WIGMORE, THE SCIENCE OF JUDICIAL PROOF (3d ed. 1937)).

2   *See generally* RICHARD EGGLESTON, EVIDENCE, PROOF AND PROBABILITY 3 (2d ed. 1983) ("It is the purpose of this book to discuss the part that probabilities play in the law, and the extent to which existing legal doctrine is compatible with the true role of probabilities in the conduct of human affairs."); Eric D. Green, *Foreword: Probability and Inference in the Law of Evidence,* 66 B.U. L. REV. 377, 377 (1986) ("[Q]uestions about the nature of proof invariably raise questions about theories of inference and the proper use of mathematical and statistical evidence and probability arguments in courts."); Symposium, *Decision and Inference in Litigation,* 13 CARDOZO L. REV. 253, 253 (1991) ("One of the more striking features of this new approach to the study of evidence was the use of symbolic notation and formal argument, particularly mathematical notation and mathematical argument."); *cf.* V.C. Ball, *The Moment of Truth: Probability Theory and Standards of Proof,* 14 VAND. L. REV. 807, 809–12 (1961) (treating frequentist theory); Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process,* 84 HARV. L. REV. 1329, 1344–50 (1971) (treating subjective theory).

3   *See generally* SHARON BERTSCH MCGRAYNE, THE THEORY THAT WOULD NOT DIE (2011) (recounting the centuries of controversy generated by Bayes' theorem).

4   Peter Tillers, *Trial by Mathematics—Reconsidered,* 10 LAW PROBABILITY & RISK 167, 169–170 (2011) (providing a nice summary of the major developments since 1970). *But see* Roger C. Park et al., *Bayes Wars Redivivus—An Exchange,* 8 INT'L COM-

debates have left unsolved some troubling problems and paradoxes in our law on proof.

The "New Logic"

One specific diagnosis of this shortcoming is that the new evidence tended to neglect the contemporaneous advances in logic.[5] The new, so-called nonclassical logic looks and sounds much like standard logic but refuses to accept some critical assumptions.[6] Most commonly, the assumption rejected is that every proposition must either be true or be false, an assumption called the principle of bivalence. But if propositions are not bivalent, so that both P and not$P$ can be true and false to a degree, then one can show that sometimes P equals not$P$—which is a rather disquieting contradiction.[7] Fashioning the new logic thus faced some challenges in its development.

The first move in the new logic of special interest to lawyers relates to and builds on the branch of modern philosophy, beginning with Bertrand Russell's work, that struggled with the problem of vagueness.[8] Work on vagueness addresses matters such as the famed sorites paradox of ancient Greece ("sorites" comes from the Greek word for heap):

MENT. ON EVIDENCE iss. 1, art. 1 (2010) (presenting an electronic exchange amongst evidence scholars debating the major issues in evidence law).

5   *See* Lea Brilmayer, *Second–Order Evidence and Bayesian Logic*, 66 B.U. L. REV. 673, 688–91 (1986) (suggesting that diagnosis); Tillers, *supra* note 4, at 171 (presenting a similar argument). Some Bayesians, however, were sympathetic to the new logic. *See, e.g.*, David A. Schum, *Probability and the Processes of Discovery, Proof, and Choice*, 66 B.U. L. REV. 825, 847–53, 865–69 (1986).

6   *See* THEODORE SIDER, LOGIC FOR PHILOSOPHY 72–73 (2010).

7   *See* Peter Suber, *Non–Contradiction and Excluded Middle*, http://www.earlham. edu/~peters/courses/logsys/pnc-pem.htm (last visited Jan. 15, 2013).

8   Bertrand Russell, *Vagueness*, 1 AUSTRALASIAN J. PSYCHOL. & PHIL. 84 (1923); *see* Bertrand Russell, *The Philosophy of Logical Atomism, in* LOGIC AND KNOWLEDGE 175, 180 (Robert Charles Marsh ed., 1956) ("Everything is vague to a degree you do not realize till you have tried to make it precise, and everything precise is so remote from everything that we normally think, that you cannot for a moment suppose that is what we really mean when we say what we think."); *see also, e.g.*, TIMOTHY WILLIAMSON, *Preface* to VAGUENESS, at xi (1996) ("[V]agueness consists in our ignorance of the sharp boundaries of our concepts, and therefore requires no revision of standard logic."); Hartry Field, *No Fact of the Matter*, 81 AUSTRALASIAN J. PHIL. 457 (2003) (countering the Williamson view); Hartry Field, *Indeterminacy, Degree of Belief, and Excluded Middle*, 34 NOÛS 1, 20 (2000) (referencing work on belief functions); Hartry Field, *Vagueness, Partial Belief, and Logic, in* MEANINGS AND OTHER THINGS (G. Ostertag ed., forthcoming 2013), *available at* http://philosophy.fas.nyu.edu/docs/IO/1158/schiffer2004b.pdf (incorporating ideas similar to fuzzy logic).

>Premise 1: if you start with a billion grains of sand, you have a heap of sand.

>Premise 2: if you remove a single grain, you still have a heap.

If you repeat the removal again and again until you have one grain of sand left, then you will by logic still have a heap. But there is no heap. Thus, heap equals nonheap. Two true premises yield an absurd conclusion (or—to picture the paradox in another common way—start with Tom Cruise's full head of hair, and begin plucking hairs, yet Tom will by logic never become bald).

At some point the heap undeniably became a nonheap. Was there a fixed boundary? No, this is not a way out—at least according to most philosophers. A different path taken in the attempt to avoid the paradox leads to the embrace of *many-valued logic*.[9] This form of logic boldly declines the simplification offered by two-valued, or bivalent, logic built on a foundation of true/false with an excluded middle. It instead recognizes partial truths. Both a statement and its opposite can be true to a degree. In other words, sometimes you have neither a heap nor a nonheap, but something that falls in between, with the statement "this is a heap" being both true and not true.[10]

The second interesting elaboration of the new logic involves developments in the field of *imprecise probability*.[11] This field of mathematics provides a useful extension of probability theory whenever

---

9   *See generally* J.C. BEALL & BAS C. VAN FRAASSEN, POSSIBILITIES AND PARADOX (2003) (introducing the concept of many-valued logic).

10   The new logic gives a pretty good answer to the sorites paradox itself, by allowing a response to "heap vel non?" in terms of a degree of heapness rather than a yes-or-no response. *See* BART KOSKO, FUZZY THINKING 94–97 (1993). But it is not a perfect answer, say the super-sophisticated. *See* R. M. Sainsbury & Timothy Williamson, *Sorites*, *in* A COMPANION TO THE PHILOSOPHY OF LANGUAGE 458, 475–77 (Bob Hale & Crispin Wright eds., 1997) ("It does not do justice to higher-order vagueness[.]"); Nicholas J.J. Smith, *Fuzzy Logic and Higher–Order Vagueness*, *in* LOGICAL MODELS OF REASONING WITH VAGUE INFORMATION 1, 1 (Petr Cintula et al. eds., forthcoming 2012), *available at* http://www-personal.usyd.edu.au/~njjsmith/papers/SmithFuzLogHOV ag.pdf ("[T]heories of vagueness based on fuzzy logic . . . give rise to a problem of *higher-order vagueness* or *artificial precision*."). *But see* TIMOTHY A. O. ENDICOTT, VAGUENESS IN LAW 77–136 (2000) ("I conclude that higher-order vagueness is *truculent*: a theory should neither deny it, nor assert a particular number of orders of vagueness, nor even assert that ordinary vague expressions are vague at all orders."). *See generally* LIARS AND HEAPS (JC Beall ed., 2003) (examining soritical paradoxes in detail); Kevin M. Clermont, *Foreword: Why Comparative Civil Procedure?*, *in* KUO–CHANG HUANG, INTRODUCING DISCOVERY INTO CIVIL LAW, at xviii n.50 (2003) (discussing self-contradictory statements).

11   *See generally* Peter Walley, *Statistical Reasoning with Imprecise Probabilities*, 42 MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY (1991) (examining the concept of, and methods of assessing, imprecise probabilities).

information is conflicting or scarce. The approach can work with many-valued logic as well as with two-valued logic. The basic idea is to use interval specifications of probability, with a lower and an upper probability. Despite its name, imprecise probability is more complete and accurate than precise probability in the real world where probabilistic imprecision prevails. In fact, traditional bivalent probability (within which I include the doctrine of random chance as well as the much newer subjective probability) appears as a special case in this theory. The rules associated with traditional probability, except those based on assuming an excluded middle, carry over to imprecise probability.

All this logic may be new, but it has an extended history of forerunners. Threads of many-valued logic have troubled thinkers since before Aristotle embraced bivalence,[12] even if their thoughts found more receptive soil in the East than in the West.[13] Imprecise probability goes back to the nineteenth century.[14] Nevertheless, the new logic has enjoyed a recent flowering, inspired by the development of quantum mechanics and instructed by those just-described advances in philosophy and mathematics.

"Fuzzy Logic"

The particular bloom known as fuzzy logic finds its roots in the seminal 1965 article by Berkeley Professor Lotfi Zadeh.[15] His critical contribution was to use degrees of membership in a *fuzzy* set running from 1 to 0, in place of strict membership in a *crisp* set classified as yes/no or as either 1 or 0. Yet fuzzy logic is not at all a fuzzy idea.[16] It became a formal system of logic, one that is by now highly developed and hence rather complicated.[17]

I do not mean to suggest that fuzzy logic resolves all the philosophical problems of vagueness (or that it is especially popular with

---

12   *See* James F. Brulé, *Fuzzy Systems—A Tutorial*, http://www.austinlinks.com/Fuzzy/tutorial.html (last visited Jan. 15, 2013) (recounting the history briefly).

13   *See* KOSKO, *supra* note 10, at 69–78 (giving an impassioned attack on the West's hostility to many-valued logic, but also fleshing out the historical account).

14   *See* GEORGE BOOLE, AN INVESTIGATION OF THE LAWS OF THOUGHT (London, Walton and Maberly 1854) (doing the early work); *cf.* JOHN MAYNARD KEYNES, A TREATISE ON PROBABILITY (1921).

15   L.A. Zadeh, *Fuzzy Sets*, 8 INFO. & CONTROL 338 (1965).

16   Zadeh substituted the term "fuzzy" for "vague" more to be provocative than to be descriptive. *See* KOSKO, *supra* note 10, at 19–20, 142, 145, 148.

17   For an accessible introduction, see Timothy J. Ross & W. Jerry Parkinson, *Fuzzy Set Theory, Fuzzy Logic, and Fuzzy Systems, in* FUZZY LOGIC AND PROBABILITY APPLICATIONS 29 (Timothy J. Ross et al. eds., 2002).

pure philosophers). I am suggesting that fuzzy logic is a very useful tool for some purposes. Of course, it has become so well-known and dominant because of its countless practical applications, especially in the computer business and consumer electronics.[18] But its theory is wonderfully broad, extending easily to degrees of truth. It thereby proves very adaptable in imaging truth just as the law does. Indeed, of the various models for handling uncertainty, fuzzy logic seems to capture best the kinds of uncertainty that most bedevil law.[19] Accordingly, writers have previously voiced suspicions that it might relate to legal standards of proof.[20]

Herein, fuzzy logic will provide a handle on how to represent our legal understandings of likelihood.[21] But it is not an exclusive tool. "In order to treat different aspects of the same problems, we must therefore apply various theories related to the imprecision of knowledge."[22] Another, compatible theory will function herein as a description of how to make decisions based on those likelihood understandings.

"Belief Functions"

Useful for that purpose is Rutgers Professor Glenn Shafer's imposing elaboration of imprecise probability from 1976.[23] His work

---

18   *See* KOSKO, *supra* note 10, at 157–200 (describing fuzzy computer systems).

19   *See* Liu Sifeng, Jeffrey Forrest & Yang Yingjie, *A Brief Introduction to Grey Systems Theory, in* 2011 IEEE INTERNATIONAL CONFERENCE ON GREY SYSTEMS AND INTELLIGENT SERVICES 1, 6 (2011).

20   *See, e.g.*, Kevin M. Clermont, *Procedure's Magical Number Three: Psychological Bases for Standards of Decision*, 72 CORNELL L. REV. 1115, 1122 n.36 (1987); Schum, *supra* note 5, at 865–69.

21   In statistical terminology, "likelihood" (the chance that the data would be observed, given a hypothesis as true) is not wholly equivalent to "probability" (the chance that a hypothesis is true, given the observed data). *See* Richard M. Royall, *Statistical Evidence*, 71 MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY, at 5–6, 28 (1997). But for most people, likelihood means probability. I use likelihood here in that way, with perhaps the connotation of an intuitive measure of probability and with the benefit of conforming to the common legal, and probabilistic, usage of "more likely than not."

22   MIRCEA REGHIS & EUGENE ROVENTA, CLASSICAL AND FUZZY CONCEPTS IN MATHEMATICAL LOGIC AND APPLICATIONS 354 (1998) (referencing Shafer's work on imperfect reasoning); *see* DAVID A. SCHUM, THE EVIDENTIAL FOUNDATIONS OF PROBABILISTIC REASONING 41, 201 (1994) ("[N]o single view of probabilistic reasoning captures all of the behavioral richness evident in such tasks.").

23   GLENN SHAFER, A MATHEMATICAL THEORY OF EVIDENCE (1976) (using "evidence" in a much broader sense than legal evidence); *see also* Glenn Shafer, *Perspectives on the Theory and Practice of Belief Functions*, 4 INT'L J. APPROXIMATE REASONING 323 (1990) (complementing his own earlier work); Lotfi A. Zadeh, Book Review, AI MAG.,

on belief functions built a bridge between fuzzy logic and traditional probability, and in the process nicely captured our legal decision-making scheme.[24] He used the word "belief" to invoke neither firm knowledge nor some squishy personal feeling, but rather the fact-finders' attempt to express their degree of certainty about the state of the real world as represented by the evidence put before them.[25] By allowing for representation of ignorance and indeterminacy of the evidence, he enabled beliefs to express uncertainty, again on a scale running from 1 to 0.[26] Indeed, his theory of belief functions rests on a highly rigorous mathematical base, managing to get quite close to achieving a unified theory of uncertainty.[27]

Belief function theory does not constitute a system of logic, unlike fuzzy logic. Instead, it is a branch of mathematics, like traditional probability.[28] Just as probability serves two-valued logic by handling a kind of uncertainty that the underlying logic system does not otherwise account for, belief function theory delivers mathematical notions that can extend many-valued logic. While probability treats first-order uncertainty about the existence of a fact, belief function notions supplement fuzzy logic by capturing and expressing the indeterminacy resulting from scarce information or conflictive evidence concerning the fact. Shafer's theory is thus similar to a scheme of second-order probability,[29] which admittedly has hitherto failed both

---

Fall 1984, at 81, 83 (reviewing SHAFER, *supra*, and treating Shafer's theory as a version of fuzzy logic's possibility theory).

24   *See* Ron A. Shapira, *Economic Analysis of the Law of Evidence: A Caveat*, 19 CARDOZO L. REV. 1607, 1614 (1998) ("In the legally relevant literature, it was Professor Glenn Shafer who introduced fuzzy measures as appropriate formalizations of epistemic functions.").

25   *See* Glenn Shafer, *The Construction of Probability Arguments*, 66 B.U. L. REV. 799, 801–04 (1986). *But cf.* DAVID CHRISTENSEN, PUTTING LOGIC IN ITS PLACE 12–13, 69 (2004) (saying that some use "belief" as an unqualified assertion of an all-or-nothing state of belief); L. Jonathan Cohen, *Should a Jury Say What It Believes or What It Accepts?*, 13 CARDOZO L. REV. 465 (1991) (using "belief," for his purposes, in the sense of a passive feeling). One could view belief as a black box, thereby avoiding my preliminary foray into fuzzy logic by starting analysis at the post-belief stage: an anti-probabilist might say that the fact-finder somehow forms a belief and that the law's concern lies in how the fact-finder should handle that belief. However, I view degrees of belief as resting on degrees of certainty, thus necessitating initial consideration of the various models for handling uncertainty.

26   For an accessible introduction, see SCHUM, *supra* note 22, at 222–43.

27   *See* Didier Dubois & Henri Prade, A Unified View of Uncertainty Theories (Mar. 7, 2012) (unpublished manuscript).

28   *See* IRVING M. COPI ET AL., INTRODUCTION TO LOGIC ch. 14 (14th ed. 2011).

29   *Cf.* Schum, *supra* note 5, at 868 (distinguishing belief functions from second-order probability).

the statistically minded[30] and psychologically minded[31] as a way to explain standards of proof.

Relation to Law

My thesis is that a better explanation of what the law does with proof lies in the new logic than in two-valued logic and its traditional probability. The explanation is indeed so good that one must entertain the notion that the law assumed and embraced the tenets of the new logic long before logicians caught up with the law.

Such an embrace by the law would not be that surprising. Law was one of the first of society's endeavors in which things appeared as neither completely true nor completely untrue. Aristotelian two-valued logic did not work for such partial truths. The common law seemed, early and intuitively, to draw many-valued logic from natural

---

30   A suggestive elaboration on traditional probability theories appeared in Neil B. Cohen, *Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge*, 60 N.Y.U. L. REV. 385 (1985) [hereinafter Cohen, *Confidence in Probability*]. That article advanced a new theory of standards of proof based on the statistical concept of confidence intervals. It described how sure a fact-finder is, employing not only a point estimate of probability but also a level of confidence, with sureness increasing as either component rises. But the attempt admittedly failed. *Compare* Neil B. Cohen, Commentary, *The Costs of Acceptability: Blue Buses, Agent Orange, and Aversion to Statistical Evidence*, 66 B.U. L. REV. 563, 569 (1986) (qualifying his own argument), *and* Neil B. Cohen, *Conceptualizing Proof and Calculating Probabilities: A Response to Professor Kaye*, 73 CORNELL L. REV. 78, 91–93 (1987) [hereinafter Cohen, *Conceptualizing Proof*] (conceding that confidence relates solely to the probability of avoiding false positives), *with* D.H. Kaye, Commentary, *Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt?*, 66 B.U. L. REV. 657, 667 n.22 (1986) (suggestively criticizing Neil Cohen's approach), *and* D.H. Kaye, *Apples and Oranges: Confidence Coefficients and the Burden of Persuasion*, 73 CORNELL L. REV. 54, 54, 56–58 (1987) (expanding his criticism).

31   A similarly suggestive attempt to explain standards of proof in terms of psychological confidence appeared in Christoph Engel, *Preponderance of the Evidence Versus Intime Conviction: A Behavioral Perspective on a Conflict Between American and Continental European Law*, 33 VT. L. REV. 435 (2009). That article posited that over the proof process's course, the fact-finder generates a level of confidence in the decision by considering the degree of coverage (which means the story accounts for all the evidence), coherence (which means it is internally consistent, plausible with respect to the fact-finder's world knowledge, and complete without striking gaps in expected components), and uniqueness (which means the absence of plausible alternative stories). *Id.* at 453. The fact-finder achieves this by an automatic or unconscious process. The clearer the view of the case that the chosen story delivers to the fact-finder, the more confident the fact-finder will be. *Id.* Against this level of confidence, the fact-finder would somehow apply the standard of proof. I tried to demonstrate his theory's prescriptive and descriptive failings in Kevin M. Clermont, *Standards of Proof Revisited*, 33 VT. L. REV. 469 (2009).

language and daily life. At about the same time, in the late eighteenth century, law and nascent probability theory began interacting.[32] That interaction caused the law to become more open about accepting uncertainty. However, traditional probability's inherent appeal prevented observers, unknowingly indoctrinated in classical bivalent logic, from seeing the law's deep logical underpinning for what it really was.[33]

Now the time has come to excavate law's multivalent foundation. The site chosen for excavation is the highly controversial subject of standards of proof. I refer to the subject in its broad sense, as covering anything that a court subjects to the proof process in order to establish truth. The subject includes many applications of law to fact, and also odd kinds of facts such as a prediction of an event, but for simplicity I shall usually refer to the whole as "facts."

The prevailing but contested view is that fact-finders should determine facts by this line of probabilistic reasoning: although given imperfect evidence, they should ask themselves what they think the chances are that the burdened party would be right if the truth were somehow to become known.

So, how would fuzzy logic and belief functions better explain standards of proof? The initial step in tying the new logic to law is to admit that the process of proof investigates a world that is not a two-valued world where disputed facts are either true or false. Instead, a good portion of the real world—by which, to be technically fine, I mean the world as perceived by humans and described by natural language—is a vague, imprecise, or many-valued world, where partial truths exist. Or, at the very least, we will never know whether a disputed fact is certainly true or false. So, the probability of truth is not the only relevant legal question. A second step is to recognize that the fact-finder's complexly constructed belief is the more relevant question. We are not as concerned with how certain the fact-finder is in a world of random uncertainty, as we are with the degree of truth the fact-finder constructs in a world of vague imprecision. The third step builds on the idea that the fact-finder will believe facts as true to a degree. We can speak of degrees of belief. Indeed, on the basis of incomplete, inconclusive, ambiguous, dissonant, and untrustworthy

---

32 *See* Kevin M. Clermont & Emily Sherwin, *A Comparative View of Standards of Proof*, 50 Am. J. Comp. L. 243, 256–57 (2002).

33 Otherwise, probabilistic theorizing influenced the law only lightly until it began quite recently a major assault. The first effective volleys of that assault on the law arguably were John Kaplan, *Decision Theory and the Factfinding Process*, 20 Stan. L. Rev. 1065 (1968), and Michael O. Finkelstein & William B. Fairley, *A Bayesian Approach to Identification Evidence*, 83 Harv. L. Rev. 489 (1970).

evidence, some of the fact-finder's belief should remain indeterminate. The output is not a probability, but what a logician would call a non-additive degree of belief in the fact's existence.[34] In the fourth step, the standard of proof will call on the fact-finder to weigh those degrees of belief.

The key distinction between probabilities and degrees of belief is subtle, as attested by the confusion among people discussing proof over the long years. Both systems numerically quantify uncertainty by using numbers in the unit interval [0,1]. But the distinction's consequences are not subtle. Degrees of belief handle imprecision better than traditional probability theory, and they better capture the effect of imperfect evidence. Also, abandoning probabilities opens the door to the logician's powerful tools for handling beliefs: fuzzy logic provides a way to represent imprecise views of the real world, and belief functions give the tools for translating beliefs about facts into legal decisions.

Another justification for so conceptualizing the process of proof is that several significant paradoxes of the law will melt away. Most paradoxes result from the limits on existing frames of reference and tools for analysis. Some paradoxes remain out of reach of today's comprehension. By "death of paradox" in my title, I simply mean that one can get a handle on many seeming paradoxes of the law by utilizing newly available frameworks and tools. As I say, this Article's view of proof works so well in this regard that it appears to have been all along the law's intuitive conceptualization of proof. Thus, after decades of reading, thinking, and writing about standards of proof, I feel as if I am finally beginning to understand them.

Present Project

This Article will attempt to make good on these grand claims, which I forward tentatively despite my sometimes-assertive tone. I shall sequentially discuss the four steps of fact-finding: (I) assessing evidence related to a single fact, (II) conjoining separate assessments of different facts, (III) analyzing the resultant beliefs, and (IV) applying the standard of proof. At each step, I shall weave in the relevant learning from the new logic and then show how it illuminates one of the key features of the law of proof: (1) the gradated scale of likeli-

---

34   *See* Shapira, *supra* note 24, at 1613–16 (distinguishing additive from non-additive). Consequently, a belief and a belief in its negation will most often not add to one. *See generally* Rolf Haenni, *Non–Additive Degrees of Belief, in* DEGREES OF BELIEF 121 (Franz Huber & Christoph Schmidt-Petri eds., 2009) (elaborating on this concept).

hood, (2) the conjunction paradox, (3) the burden of production, and (4) the burden of persuasion.

The primary focus of the Article is descriptive and explanatory, not prescriptive other than by implication. I seek to reveal what the law actually tells its fact-finders to do. Over the centuries the law's charge to fact-finders has evolved, by a process entailing considerations of both ideal fact-finding and also human limitations. I am not championing new ideals or new limitations. My real interest here lies in exposing the proof process that the law has chosen. I believe and shall try to demonstrate that the law has embraced what became the new logic, likely because that version of logic captures the epistemic function of law better than classical logic and probability theory.

A deeper understanding of what the law says about its process of proof could surely lead to more knowledge and to improved law. It could stimulate new research into how fact-finders actually decide. It could lead to legal reforms, such as clearer instructions on the standards of proof. But reform is not the present exploratory project. Indeed, one message of this Article might be that the law currently does a much better job in structuring fact-finding than one would guess by reading its many critical commentators.

To encapsulate, the aim of this Article is to apply the new logic to the law. The conclusion will be that this logic snaps onto the law as a seemingly perfect fit.

I.   ASSESSING EVIDENCE

This Part will convey the basics of fuzzy logic, when used as a way to gauge degrees of truth based on assessment of evidence related to a disputed fact. Then, this Part will show how the law has seemingly employed fuzzy logic to construct, for use in its standards of decision, a gradated scale of likelihood stretching across the spectrum from the slightest possibility up to virtual certainty.

A.   Theories

1.   Psychology Theories

How does the fact-finder reach a decision? No one knows. Psychologists cannot tell us exactly how people evaluate and combine evidence and, for that matter, can tell us almost nothing about how they weigh assessments of evidence against a standard of proof.[35]

---

35   *See* Clermont, *supra* note 31, at 475–79, 485 (recounting the state of psychological knowledge on evidence processing and standards of proof).

Introspection might suggest that there is a knowledge arc, leading up by induction and abduction, and then down by deductive testing.[36] The upward arc rests on observations, which generate new hypotheses explaining the observations. The downward arc involves testing the hypotheses to reach conclusions. Throughout, there is evaluation, combination, and weighing of evidence by some method.

To expose that method, social scientists have tried to model the cognitive black box by experimentally comparing inputs and outputs. One result is the model called *information integration theory*.[37] It tries to describe how humans naturally evaluate and combine information to produce judgment. Although only one of many contesting theories, and a relatively optimistic one at that, information integration theory has suggestive powers, making description worthwhile. According to the theory, the human decision-maker who has to make a finding on a fact's existence would begin with an initial impression, or predisposition, and then would process additional pieces of information. Each of these, including the initial impression, would receive a scale value, which is a measure of the likelihood of the fact's existence. Each would also receive a weighting factor, which is a measure of evidential importance that takes into account both directness and credibility. The decision-maker would then combine these into a weighted average that determines the fact's existence.

Even if some such theory reflects reality, it is clear that humans do not naturally use perfectly "rational" techniques, but instead use less accurate "intuitive" techniques.[38] The employed techniques are also subject to all sorts of heuristics and other biases.[39] Any account would further have to incorporate the metaphorically dual processes, automatic and systematic, of cognitive processing.[40]

---

36   *See* David A. Schum, *A Science of Evidence: Contributions from Law and Probability*, 8 LAW PROBABILITY & RISK 197, 203–04 (2009) (crediting DAVID OLDROYD, THE ARCH OF KNOWLEDGE (1986), for this image).

37   *See* Martin F. Kaplan, *Cognitive Processes in the Individual Juror, in* THE PSYCHOLOGY OF THE COURTROOM 197 (Norbert L. Kerr & Robert M. Bray eds., 1982); *see also* Jennifer Groscup & Jennifer Tallon, *Theoretical Models of Jury Decision–Making, in* JURY PSYCHOLOGY 41 (Joel D. Lieberman & Daniel A. Krauss eds., 2009) (sketching some contesting models); Nancy Pennington & Reid Hastie, *Juror Decision–Making Models: The Generalization Gap*, 89 PSYCHOL. BULL. 246 (1981) (evaluating various models). For treatment of the story model, see *infra* text accompanying note 116.

38   *See generally* DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011). For treatment of humans' bounded rationality, *see infra* text accompanying note 84.

39   *See* J. Alexander Tanford & Sarah Tanford, *Better Trials Through Science: A Defense of Psychologist–Lawyer Collaboration*, 66 N.C. L. REV. 741, 748–59 (1988).

40   *See* Anna Ronkainen, *Dual–Process Cognition and Legal Reasoning, in* ARGUMENTATION 2011, at 1, 1 (Michal Araszkiewicz et al. eds., 2011), *available at* http://ssrn.com/

In actuality, then, the fact-finders' performance of the cognitive process would usually be approximate and non-quantitative. The fact-finders would take a stab at assessing all the related evidence, perhaps by some method similar to information integration. Then, any judged likelihood might find expression in terms of a limited set of broad categories such as more likely than not, high probability, and almost certainty.

2. Probability Theories

A partly separate question from what fact-finders can do, actually do, or think they are doing, on the one hand, is what the law should or does tell the fact-finder to do, on the other hand. Surprisingly, even mathematicians cannot agree on how a fact-finder should perform the task of processing evidence, and so cannot unanimously guide the law on an ideal path to take.[41] The most popular candidate for rigorously evaluating and combining related evidence is the Bayesian approach, utilizing subjective probabilities.

Subjective probability theory allows us to speak of the likelihood of a single event. A subjective probability measures an individual's personal judgment about how likely a particular event is to occur or has occurred. The theory is

> based on the notion that it makes sense to ask someone what he would do if offered a reward for guessing correctly whether any proposition, designated X, is true or false. If he guesses that X is true under these circumstances, we say that *for him* the subjective probability of X, written P(X), exceeds fifty percent. Symbolically, P(X) > .5. If he would be equally satisfied guessing either way, then we say that, for him, P(X) = .5.[42]

Upon expanding the measure into a complete scale of probabilities from 0 to 1 and postulating the usual logical operators, subjective probabilities follow most of the rules of frequentist probabilities.[43]

---

abstract=2004336 ("The dual-process framework is a set of theories on human cognition in which cognition is seen as consisting of (at least) two substantially different yet interdependent systems: the older, faster, partly unconscious and automatic System 1 and the newer, slower, fully conscious and considered System 2.").

41   *See* Peter Tillers, *Introduction: A Personal Perspective on "Artificial Intelligence and Judicial Proof,"* 22 CARDOZO L. REV. 1365 (2001) (discussing different mathematical approaches to judicial proof).

42   Tribe, *supra* note 2, at 1347 (citing the innovative work of LEONARD J. SAVAGE, FOUNDATIONS OF STATISTICS (1954)).

43   *See id.* at 1347–48. *Compare id.* at 1348 n.63 (accepting the product rule because he is assuming bivalence), *with infra* text accompanying note 110 (rejecting the product rule for subjective probabilities in fact-finding).

Bayes' theorem links the perceived probability before and after observing evidence. The starting point is P(A), the prior probability of A. Then the posterior probability of A, after accounting for evidence B, is P(A|B); this is a so-called conditional probability, which may be read as the probability that A will occur if B is known certainly to have occurred. P(A|B) calculates to be P(A) multiplied by the support B provides for A, a support that Thomas Bayes (or really Pierre Simon Laplace) equated to P(B|A) / P(B).

Despite its internal mathematical soundness, and despite the many insights it generates for law, many observers from various disciplines have serious doubts about whether Bayes' theorem should be seen to play a broad role in legal fact-finding.[44] First, its prior probability comes out of thin air, as some sort of subjective guess. The more objective supposition of 50%, on the thought that the fact is either true or false, comports neither with reality nor with where the law tells the fact-finder to begin.[45] Second, Bayes' theorem leaves no place for indeterminacy, thus painting the world as black and white even though most of the world appears in shades of gray. It accordingly does not handle well the situation of conflicting or scarce information, using a fudge factor to account for the state of the evidence.[46] Third, its mathematical approach is not realistic, of course. It does not conform to the way intuitive humans arrive at prior probabilities or the way they combine them with new evidence to produce posterior probabilities.[47]

3.   Zadeh's Fuzzy Logic

a.   Basics

Fuzzy logic envisages degrees of membership in a so-called fuzzy set, with membership valued anywhere between 0 and 1. That is, x's membership in a fuzzy set H of the universe X may take values throughout the whole interval [0,1], rather than just the two values of

---

44   *See, e.g.*, Pennington & Hastie, *supra* note 37, at 262–68; Shafer, *supra* note 25, at 809–16. *Compare* Paul Bergman & Al Moore, *Mistrial by Likelihood Ratio: Bayesian Analysis Meets the F-Word*, 13 CARDOZO L. REV. 589, 590 (1991) (attacking), *with* D.H. Kaye, Commentary, *Credal Probability*, 13 CARDOZO L. REV. 647 (1991) (defending).

45   *See* Leonard R. Jaffee, *Of Probativity and Probability: Statistics, Scientific Evidence, and the Calculus of Chance at Trial*, 46 U. PITT. L. REV. 925, 980–85 (1985).

46   *See* Lea Brilmayer & Lewis Kornhauser, *Review: Quantitative Methods and Legal Decisions*, 46 U. CHI. L. REV. 116, 135–48 (1978).

47   *Compare* SAMUEL KOTZ & DONNA F. STROUP, EDUCATED GUESSING (1983) (discussing the sophisticated techniques of theory), *with* Craig R. Callen, *Notes on a Grand Illusion: Some Limits on the Use of Bayesian Theory in Evidence Law*, 57 IND. L.J. 1 (1982) (discussing the simplified practices of law).

0 and 1. This range allows a more complete and accurate expression of membership, when membership is imprecise.

Take as an example the set A of men from five to seven feet tall. This would be a so-called crisp set. Membership of element x in set A, represented by $\chi_A(x)$, is not vague. The values of χ_A can thus be only either 0 or 1, at least if we ignore complications at the nanoscale level.

Contrast the set H of men somewhere near six feet. It is a fuzzy set. Membership of element x in H, represented by $\mu_H(x)$, is imprecise. The values of μ_H may start at 0 for a tiny person, but they soon increase by some function to a value of 1 at precisely six feet, and then start decreasing. The membership function could be linear, but it can take on any shape as appropriate.

So Tom might be completely in set A but have a degree of membership in H of .5. The following figure represents these two sets, with (a) representing the crisp set of men from five to seven feet tall and (b) being one representation of the fuzzy set of men somewhere near six feet tall:[48]
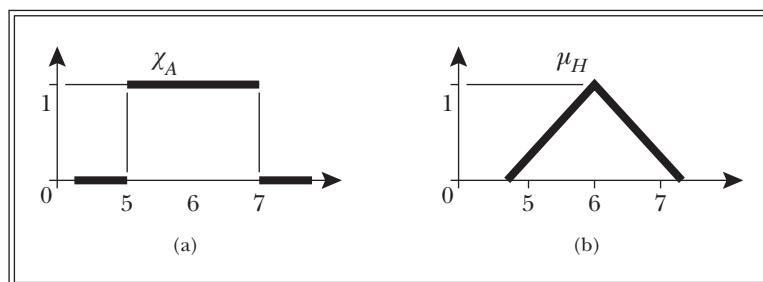
b.   Linguistics

Important to note is the role of qualifying language in the prior example as a means of expressing membership. *Evaluative linguistic expressions* are words like small, medium, and big; another example is about, roughly, or near, when used in contrast to not-at-all or really. These words may not be a large part of our natural language. Yet, they are an important part of that language. They do a lot of work. People use them all the time to evaluate a thing or situation and to communicate their evaluation. People thereafter use them for classification, decision-making, and other tasks.

People employ *linguistic hedges* to modify their evaluations further. Words such as very or extremely, and fairly or almost, are examples. These words allow people to create a gradated scale for their

---

48   The figure comes from Ross & Parkinson, *supra* note 17, at 30.

evaluations. The following figure represents a scale for size, ranging from v_L as the left bound of values that are very small, through medium, and on to v_R as the right bound of values that are very big:[49]
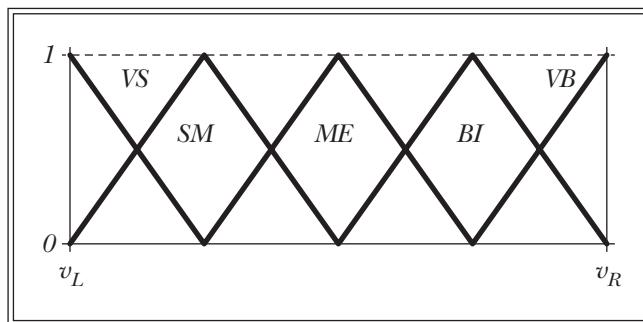
This scale actually consists of three overlapping fuzzy subsets based on the evaluative linguistic expressions of small, medium, and big. Linguistic hedges subdivide the three subsets as suggested by the four smaller triangles at the top, perhaps equating to very small, almost medium, almost big, and very big. The result is seven gradations of meaning.

Obviously, these seven categories and labels are imprecise. But they are supposed to be, and they work quite well in real life. Fuzzy logic better comports to our view of the real world than does classical bivalent logic with its excluded middle. In the real world, boundaries of things often appear indistinct, thus making important or even essential the human ability to process fuzzy information.

Some research suggests that such verbal categories work better than numerical translations.[50] Other research indicates that they work best as a matter of natural language if the categories are equally

---

49 The figure comes from Vilém Novák, *Modeling with Words*, SCHOLARPEDIA, http://www.scholarpedia.org/article/Modeling_with_words (last visited Jan. 15, 2013).

50 *See* Alf C. Zimmer, *Verbal vs. Numerical Processing of Subjective Probabilities, in* DECISION MAKING UNDER UNCERTAINTY 159, 180 (Roland W. Scholz ed., 1983). On the one hand, considerable empirical work on the legal standards of proof suggests that fact-finders show considerable confusion in translating the standards into numerical probabilities. The results argue for a verbal approach to standards of proof. *See* Clermont, *supra* note 20, at 1144–50 (recounting the empirical studies that show fact-finders' difficulties in comprehending standards of proof). On the other hand, not only is such translation unnecessary, but also it might be a wrongheaded step if fuzzy logic is really in use. The prior empirical work might therefore need redoing. *See* Mandeep K. Dhami, *On Measuring Quantitative Interpretations of Reasonable Doubt*, 14 J. EXPERIMENTAL PSYCHOL.: APPLIED 353, 362 (2008) (developing a new technique for empirical work on standards of proof, called the membership function and based on

sized and shaped.[51] It might even be feasible to construct and employ an ordinally ranked vocabulary of likelihood that improves effective interpersonal communication.[52]

c.   Evaluating and Combining Evidence

Fuzzy logic has formal rules for combining likelihoods,[53] once evaluated by one method or another.[54] Moreover, to jump ahead to belief functions, the theoretical work thereon consists mainly of developing tools for combining pieces of evidence to determine a likelihood. In particular, its very prominent Dempster-Shafer rule governs the task.[55] That rule is very complicated because it abstractly addresses the problem in the most general terms possible (Bayes' the-

---

fuzzy logic, and highlighting "the need to reevaluate the reliability and validity of past research findings on quantifying" standards of proof).

   51   *See* Zimmer, *supra* note 50, at 166.

   52   *See* Mandeep K. Dhami & Thomas S. Wallsten, *Interpersonal Comparison of Subjective Probabilities: Toward Translating Linguistic Probabilities*, 33 MEMORY & COGNITION 1057 (2005).

   53   *See* Petr Hajek, *Fuzzy Logic*, *in* THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (Edward N. Zalta ed., 2010), *available at* http://plato.stanford.edu/archives/fall 2010/entries/logic-fuzzy (discussing the varying versions of fuzzy logic).

   54   *Cf.* Brulé, *supra* note 12 ("The skeptical observer will note that the assignment of values to linguistic meanings (such as 0.90 to 'very') and vice versa, is a most imprecise operation. Fuzzy systems, it should be noted, lay no claim to establishing a formal procedure for assignments at this level; in fact, the only argument for a particular assignment is its intuitive strength. What fuzzy logic does propose is to establish a formal method of operating on these values, once the primitives have been established.").

   55   *See* SHAFER, *supra* note 23, at 6, 57–67 (using orthogonal sums); Jeffrey A. Barnett, *Computational Methods for* A Mathematical Theory of Evidence, *in* CLASSIC WORKS OF THE DEMPSTER–SHAFER THEORY OF BELIEF FUNCTIONS 197, 198–204 (Ronald R. Yager & Liping Liu eds., 2008). By the Dempster-Shafer rule,

> we construct a belief function to represent the new evidence and combine it with our "prior" belief function—i.e., with the belief function that represents our prior opinions. This method deals symmetrically with the new evidence and the old evidence on which our prior opinions are based: both bodies of evidence are represented by belief functions, and the result of the combination does not depend on which evidence is the old and which is the new.

SHAFER, *supra* note 23, at 25.

orem turns out to be a special case of that approach).[56] The rule is also quite contested, generating many competitors.[57]

In the end, these formal approaches are clearly not realistic representations of human cognition.[58] Humans do not naturally use rational techniques like the Dempster-Shafer rule any more than they calculate Bayes' theorem; they instead use intuitive techniques in a non-quantitative and approximate fashion. Consequently, the law has generally left the combination of related evidence evaluations to its fact-finders' instinctive treatment.

Fortunately, fuzzy logic is not dogmatic on the method used to evaluate or combine pieces of evidence that reinforce or undermine each other. It is compatible with the fact-finders' combining all the related evidence by any means. Just one possibility would be information integration's weighted average, described above.[59] In fact, a weighted-average approach commonly appears in the decision-making process of today's fuzzy computer programs.[60]

My interest in this Article is not so much the initial eyeing of evidence, but rather the subsequent steps that build to a standard of proof's application. For my purposes, the new logic mainly illuminates a way for people to express their views of the evidence, prior to handling those views according to established rules of reasoning such as conjunction. That is, fuzzy logic in this broad sense gives a new and effective way to explore humans' expression of their assessment.[61] And it then opens the door to subjecting those expressions to so-called approximate reasoning, which despite its name can be quite rigorously performed.[62]

---

56   For a comparison of Bayesian probability judgments and belief functions, see Glenn Shafer & Amos Tversky, *Languages and Designs for Probability Judgment, in* CLASSIC WORKS OF THE DEMPSTER–SHAFER THEORY OF BELIEF FUNCTIONS, *supra* note 55, at 345.

57   *See* Kari Sentz & Scott Ferson, Combination of Evidence in Dempster–Shafer Theory 17–27 (Apr. 2002) (unpublished manuscript), *available at* http://www.sandia.gov/epistemic/Reports/SAND2002-0835.pdf (describing thirteen alternatives).

58   *See* Schum, *supra* note 5, at 852–53 (recounting the lack of empirical support).

59   *See supra* text accompanying note 37.

60   *See* KOSKO, *supra* note 10, at 176–80 (describing its use in fuzzy computer systems).

61   *See* Novák, *supra* note 49 ("Mathematical fuzzy logic has two branches: *fuzzy logic in narrow sense* (FLn) and *fuzzy logic in broader sense* (FLb). FLn is a formal fuzzy logic which is a special many-valued logic generalizing classical mathematical logic. . . . FLb is an extension of FLn which aims at developing a *formal theory of human reasoning*.").

62   *See* Ronald R. Yager, *New Paradigms for Reasoning with Uncertain Information,* 13 CARDOZO L. REV. 1005, 1017–24 (1991) (explaining approximate reasoning).

I therefore will not pursue fuzzy logic's detailed rules, which constitute the new logic in a narrow sense, nor will I weigh the disputes over the rules' details. Naturally enough, the law built on the intuition that partial truths exist, but it never adopted all the formal complications that logicians have refined since 1965.

d. Fuzziness versus Probability as a Means of Expressing Assessment

I propose considering the broad version of fuzzy logic as the legal model for human expression of uncertainty, in preference to assuming a probability-based bivalent view. The significance of adopting it for this purpose will become evident upon distinguishing a fuzzy statement from a traditionally probabilistic one.[63]

A few anticipatory words of qualification are in order. I am not an anti-probabilist. I am not arguing against a probabilistic approach if "probabilistic" carries its capacious sense of accepting uncertainty and mathematically accounting for it. I am arguing against traditional probability theory when appended to a bivalent view of the world. What I am proposing is a nontraditional means for expressing uncertainty.

Describing different uncertainties—Both a traditionally probabilistic description and a fuzzy one can be accurate statements, but they describe different states. The key distinction is that probability here depends on the existence of a crisply bivalent world. Fuzzy logic accepts an imprecisely multivalent world.

On the one hand, fuzziness is a way, for example, to describe *event imprecision.* It measures what has occurred—actually, the degree to which the event occurred—which can be vague. On the other hand, probability is a way to describe *event occurrence.* It can measure the chance that the event will occur or not.

The probability of whether an event occurs in a bivalent world is normally random, as contrasted with the nonrandom uncertainty of vagueness that fuzziness additionally measures. Probability expresses the chance of whether something will occur, all the while knowing it will occur or not on an all-or-nothing basis. It is a mathematical supplement to bivalent logic, used to account for that one kind of uncertainty. Fuzziness expresses vagueness as a degree of membership. It

---

63   *See generally* Mark Colyvan, *Is Probability the Only Coherent Approach to Uncertainty?*, 28 RISK ANALYSIS 645 (2008) (arguing that probability theory's dealing with uncertainty stands on controversial premises and suggesting examples of non-probabilistic uncertainty); Bart Kosko, *Fuzziness vs. Probability*, 17 INT'L J. GEN. SYS. 211 (1990) (discussing and contrasting fuzziness and probability).

builds its inclusive measure of uncertainty right into the basics of the multivalent logic system.

Probability conveys what we know, when in possession of only partial knowledge, about whether an event will occur. With more information, the uncertainty in the probability will dissipate, and if the event's occurrence becomes known the probability itself will morph into a value of 1 or 0. By contrast, fuzziness conveys all the information we have about an event, which most often ends up expressed as a partial truth. More information will increase the fuzziness of set membership, because any crisp lines become harder to maintain. That is, as one acquires more information about the world, one sees a greater need for measuring fuzziness in lieu of bivalent categorization.

Probabilism and fuzziness can describe much more than events. Maybe another example would help. If a probabilist says, "There is a 30% chance that Tom is tall," the speaker supposes that Tom is either tall or not tall, and given imperfect evidence he thinks that it is only 30% likely that Tom would end up in the tall category upon accurate measurement. But when a fuzzy logician says, "Tom's degree of membership within the set of tall men is .30," he means that Tom is not very tall at all. The difference is real and considerable. It derives from the fact that the probabilist is assuming bivalence with an excluded middle, so that one is tall or not, while the fuzzy logician is speaking of a world where one can be more or less tall.

Choosing between models—Which model to use, probability or fuzzy logic, depends on what one is trying to describe. If the fact in question is or is assumed to be nonvague, and thus readily distinguishable from its opposite, and its occurrence is subject only to random uncertainty, then probability is appropriate. For a probability example: will I pick a black ball from the urn? However, if the fact is vague, and most facts in the world are vague, fuzzy logic is the way to go. For a fuzzy example: how black is this grayish ball?

The choice between probabilism and fuzziness is important. The kind of statement one can make will depend on the choice made. "You paint one picture of the world if you say there is a 50% chance that an apple sits in the refrigerator. You paint a different picture if you say half an apple sits in the refrigerator."[64] The two models are not fully interchangeable, even though people tend to treat them so. People use probability loosely for any sort of uncertainty. They use it to express fuzziness. But it is inappropriate for that purpose.

I am coming to the choice that the law has made. But at this point, it is natural for the reader to jump to the conclusion that the

---

64 Kosko, *supra* note 10, at 15.

law in its fact-finding usually wants to know if an apple is in the refrigerator, not whether it is half eaten. The court wants to know if Tom was or was not the alleged perpetrator. Just to slow you up, however, I point out that in legal fact-finding no one is ever going to be able to look inside the refrigerator. Also, much more significantly, I can pose another example that makes it much less clear which sort of statement the law seeks. Think of a somewhat sloppily drawn circle: is it more appropriate to say (i) there is a 90% probability that it is a perfect circle or (ii) it has a .90 membership in the set of circles?[65] An analogy to the circle would be the law's trying to determine fault, when degrees of fault are the reality. But also analogous would be causation, consent, coercion, good faith, intent, and a host of other legal issues. After all, remember that Bertrand Russell saw all of natural language as vague.[66] Many, many legal issues are fuzzy concepts, in that they draw indistinct lines, most often unavoidably—and many of these fuzzy concepts are subjects of proof through litigation.[67]

So, the important choice between probabilism and fuzziness is not an easy one. Bearing on that choice, however, consider three advantages of fuzzy logic.

First, it is more accurate than probability whenever one encounters nonrandom uncertainty, such as vagueness. It picks up the extra information about vagueness, extra information expressed in natural language but lost upon classification into a crisp set. Recall that fuzziness includes the imprecision of an event, while probability describes only the chance of the event. The precision of probability thus turns out to be a vice rather than a virtue. Probability has the advantage of bivalent simplicity, but it will often be misleading in the real world of fuzziness:

> The question is not whether the glass is half empty or half full. If we had to say all or none, the question is, is the glass full or empty.

---

65   *See id.* at 44–46 (using this image).

66   *See supra* note 8.

67   *See* ENDICOTT, *supra* note 10 (arguing that vagueness plays a significant role in law, a role not owing solely to the vagueness of language); LEO KATZ, WHY THE LAW IS SO PERVERSE 139–56 (2011) (cataloging examples of vague concepts); Andrei Marmor, *Varieties of Vagueness in the Law* (USC Legal Studies Research Paper No. 12-8, Apr. 2012), *available at* http://ssrn.com/abstract=2039076 (articulating the different types of vagueness in law, beyond those entailing only a simple sorites sequence); Scott Soames, *Vagueness in the Law, in* THE ROUTLEDGE COMPANION TO PHILOSOPHY OF LAW 95 (Andrei Marmor ed., 2012) (bridging between philosophical logic and legal philosophy). Consequently, legal scholars are increasingly using fuzzy logic. *See* Michael T. Nguyen, Note, *The Myth of "Lucky" Patent Verdicts: Improving the Quality of Appellate Review by Incorporating Fuzzy Logic in Jury Verdicts*, 59 HASTINGS L.J. 1257, 1261 n.28 (2008) (listing examples).

> [Either answer is a half-truth.] . . . That is the real state of the world. We don't mean that there is a 50% probability that the glass is full. We mean half a glass. If for some cultural reason we limit what we say to the two bivalent options of all or none, true or false, yes or no, then we pay the price . . . .[68]

Second, the precision of probability becomes particularly troublesome when trying to gauge a probability resting on imperfect evidence that leaves a lot indeterminate. Fuzzy expression can better handle incomplete, inconclusive, ambiguous, dissonant, and untrustworthy evidence.[69]

Third, another advantage of fuzzy logic is that it is the more inclusive system. Many-valued logic includes two-valued logic. Being a form of many-valued logic, fuzzy logic neither requires nor forbids that anything be of an on-or-off nature, true or false, completely inside a set or outside that set. The two-valued logic of probability demands the existence of sets with strict membership classifications exclusively, an in-or-out characteristic symbolized respectively by values of either 1 or 0. But those crisp sets are a kind among fuzzy sets.

---

68 KOSKO, *supra* note 10, at 25–26; *see id.* at 33 ("At the midpoint you cannot tell a thing from its opposite, just as you cannot tell a half-empty glass from a half-full glass.").

69 *See generally* MACIEJ WYGRALAK, VAGUELY DEFINED OBJECTS (1996). The degree of membership in a fuzzy set can itself be imprecise or otherwise uncertain, making what is called an *ultra-fuzzy* set or a type-2 fuzzy set, which operates as an initial step toward type-*n* fuzzy sets and operates in contrast to the type-1 fuzzy sets discussed up to here. *See* Mark Jablonowski, *An Ultra–fuzzy Model of Aggregate Growth in Catastrophic Risk Potentials*, 2008 ANN. MEETING N. AM. FUZZY INFO. PROCESSING SOC'Y (May 19–22, 2008) (modeling using ultra-fuzzy sets). Basically, representation of the uncertainty in the degree of membership comes out in a third dimension from each degree of membership represented in two dimensions. *See* Jerry M. Mendel, *Type–2 Fuzzy Sets and Systems: An Overview*, IEEE COMPUTATIONAL INTELLIGENCE MAG., Feb. 2007, at 20, *available at* http://sipi.usc.edu/~mendel/publications/MENDEL%20CI%20Magazine%202007.pdf. The third dimension can be projected back into the two dimensions to create the footprint of uncertainty, or FOU, which I mention just to suggest how cool the terminology in this field gets.

Of course, the logical operators become much more complicated to account for this third dimension. *See* Nilesh N. Karnik & Jerry M. Mendel, *Operations on Type–2 Fuzzy Sets*, 122 FUZZY SETS & SYS. 327 (2001); *see also* Jerry M. Mendel, *Type–2 Fuzzy Sets and Systems: How to Learn About Them*, IEEE SMC ENEWSL. (Sys., Man & Cybernetics Soc'y, New York, N.Y.), June 2009, *available at* http://sipi.usc.edu/~mendel/publications/MendelSMCeNewsletter6-09.pdf (discussing existing work on ultra-fuzzy sets). However, here there is no reason to explore the sophistication of logical operators for ultra-fuzzy sets, because the law seems to treat all measures of truth simply as type-1 fuzzy sets and instead roughly accounts for other second-order-like uncertainty through belief functions.

Bivalence is a special case of multivalent logic. The world of black or white is a special, extreme case of the world shaded in grays.

In many situations, a single fact is subject both to vagueness and to occurrence uncertainty. Although probability theory can say little on how to reason about things that are not completely true or false, fuzzy logic can handle the more mixed and complex situations. For an example of what might be called "normative" uncertainty,[70] it might be that there was a .70 blameworthy act. Or while the act was completely blameworthy, there was a 70% chance that it occurred, creating "factual" uncertainty. But what about a 70% chance of a .70 degree of fault? Then, these two kinds of uncertainty need to be integrated.

Moreover, a decision may rest on a number of facts, some of which demonstrate only occurrence uncertainty, while others show some vagueness as well. The need in decision-making to combine these fact-findings counsels the use of one logic system, because "we cannot coherently countenance two different kinds of degree of belief."[71] To combine the facts, a common currency is necessary.

The inclusiveness of fuzzy logic suggests its use for each fact and for combining evaluations of different facts. Indeed, it can effortlessly express even a traditional probability as membership in a set; that is, probability is the degree to which the imagined universe of all tries would belong to the set of successful tries.[72] This compatibility is essential. It allows easy combination of a randomly uncertain set with an imprecise set, in accordance with fuzzy logical operators.[73]

In sum, fuzzy logic provides the needed common currency. It can handle all kinds of facts, and can do so much better than probability. The legal reader cannot dodge my argument by concluding that fuzziness reaches, if any, only some kinds of facts. Instead, fuzziness handles facts exhibiting random uncertainty as well as those showing vagueness, facts embodying both factual uncertainty and nor-

---

70   *See* Ariel Porat & Eric A. Posner, *Aggregation and Law*, 122 YALE L.J. 2, 2 (2012) (coining the terms "normative aggregation" and "factual aggregation").

71   Nicholas J.J. Smith, *Degree of Belief Is Expected Truth Value*, *in* CUTS AND CLOUDS 491, 491 (Richard Dietz & Sebastiano Moruzzi eds., 2010) (cataloging the difficulties that would come from entertaining both probabilities and degrees of belief, when the two underlying logic systems employ different operators).

72   *See* KOSKO, *supra* note 10, at 55–64 (calling this set "the whole in the part").

73   *See infra* Part II.A. (explaining the MIN operator). This approach would also handle the uncertainty of whether an imprecise event occurred at all. *See* Charles M. Yablon, *On the Allocation of Burdens of Proof in Corporate Law: An Essay on Fairness and Fuzzy Sets*, 13 CARDOZO L. REV. 497 (1991) (treating a transaction's fairness as a fuzzy set, while trying at the same time to account for subjective probability of occurrence).

mative uncertainty, and factual events and legal constructs. Law could choose fuzziness as its sole mode of measurement.

e.   The Law's Choice

My concern in this Article is primarily to unearth what the law actually tells its fact-finders to do. Even if in the end it remains up in the air as to which model, fuzzy logic or probability, would be the superior vehicle for the legal proof process, my real interest here is in which model the law actually chose. This is a descriptive rather than prescriptive question: which model better expresses the law's scale of likelihood of truth, which better explains the legal treatment of the well-known conjunction paradox, and which better effectuates the burdens of production and persuasion? On the descriptive question, I believe and shall try to demonstrate that the law embraces fuzzy logic.

In its instructions to fact-finders, the law does not explicitly distinguish between probabilism and fuzziness. It instead speaks in terms of a common currency, mixing the questions of whether and how much an event occurred. Whether the issue is occurrence or blameworthiness of an act, the law deals only in degrees of truth. Asking for a single measure of uncertainty makes sense only in fuzzy logic, because it accounts for the various kinds of uncertainty. Relying instead on the mathematical supplement of the probability calculus would be so awkward and incomplete as to be nonsensible. I therefore submit that the law treats all measures of truth simply as fuzzy sets.

I would further submit, if pressed, that the reason for the law's choosing to speak in terms of degrees of membership is that they behave more appropriately than probabilities in a world filled with various kinds of uncertainty.[74]

Imagine that the law is trying to determine if Tom was at fault. A number of features of this issue of fault indicate that the actual and better approach for law is fuzzy logic. First, we will never know the answer as a 1 or a 0. Therefore, we should not be worrying too much about specifying the chance of a 1 turning up. Second, ours is not a crisp world, so the law is often not interested in establishing that the truth value of an element is a 1 or 0. Instead, it wants to ascertain whether the element has a sufficient truth value for the purpose at hand. Third, any conclusion based on evidence is necessarily uncertain for five reasons: "Our evidence is never complete, is usually inconclusive, is frequently ambiguous, is commonly dissonant to some

---

74   *See* Shapira, *supra* note 24, at 1614–15 (arguing for the superiority of degrees of belief).

degree and comes to us from sources having imperfect credibility."[75] Fourth, the fact-finder might entertain thoughts of both randomness and vagueness, that is, both a sense that Tom was .70 likely to have been completely at fault and also that Tom was at fault to a .70 degree. Fifth, given that some issues in a case might demand the more inclusive measure of imprecision, logical coherency requires that the same type of measure apply to every issue.

I am not calling for a major shift in conceiving the standards of proof. After all, fuzzy logic is not antithetical to classical logic. All I am saying is that the law appreciates that more kinds of uncertainty than that of event occurrence are at play. The law therefore uses a logic appropriate to the task. Fuzzy logic moves its measure of uncertainties into the basics of the system, rather than leaving their treatment to some sort of afterthought. That non-radical move does not call for an overhaul of legal language or imagery, and it even makes many legal consequences easier to comprehend.

Indeed, I am not even saying that if we recognize the role of multivalence, we need to abandon the probability idiom. It can be quite expressive in the realm of standards of proof. Nonetheless, when plunging to the depths, we need always to remember that the legal foundation is ultimately fuzzy in nature. Here is a last clever image to make the point as to what the law deals in:

> Suppose you had been in the desert for a week without drink and you came upon two bottles marked K and M [and marked, respectively, with a .91 membership in the fuzzy set of potable liquids and a .91 probability of being a potable liquid]. Confronted with this pair of bottles, and given that you must drink from the one that you chose, which would you choose to drink from? Most people, when presented with this experiment, immediately see that while K could contain, say, swamp water, it would not (discounting the possibility of a Machiavellian fuzzy modeler) contain liquids such as hydrochloric acid. That is, membership of 0.91 means that the contents of K are fairly similar to perfectly potable liquids, e.g. pure water. On the other hand, the probability that M is potable "equals 0.91" means that over a long run of experiments, the contents of M are expected to be potable in about 91% of the trials. In the other 9% the contents will be deadly—about 1 chance in 10. Thus, most subjects will opt for a chance to drink swamp water.[76]

---

75   Schum, *supra* note 36, at 216 (internal quotation marks omitted).

76   Bogdan R. Kosanovic, Fuzziness and Probability 2–3 (Feb. 8, 1995) (unpublished manuscript) (emphasis omitted), *available at* http://www.neuronet.pitt.edu/ ~bogdan/research/fuzzy/fvsp/fvsp.html. In early 2012 some congressional offices received threatening letters containing a suspicious powder. The letter promised additional mailings and said there was a "'10 percent chance that you have just been

B. Legal Application: Gradated Likelihood

The way to apply fuzzy logic to fact-finding in the legal system is to envisage the fuzzy set of true facts and ask for *x*, as a particular element of a claim or defense, what is its degree of membership in that set. Recall that membership represents how much a variable is in the set. The membership here will be partial. It will tell how true the fact-finder finds the element to be. One could express $\mu(x)$ as truth(x). Membership thereby creates degrees of truth.[77]

While this membership will turn on likelihood of truth in a sense, it is in a sense different from the classical understanding of the fact-finder's subjective probability that the element is true. Such subjective probability crisply deals with the probability of *x* being in actuality 1, while fuzzy logic vaguely deals with a degree of truth. The degrees of truth range from 0 to 1 in fuzzy theory, but in practice they find expression most often as words, or evaluative linguistic variables that use linguistic hedges to cover all the intervals of partial truth between completely false and completely true.

Fuzzy logic's schema well describes the law's scale of likelihood that I have previously documented.[78] For a significant example, the law today limits its choice to no more than three standards of proof—preponderance, clearly convincing, and beyond a reasonable doubt—from among the infinite range of probabilities stretching from slightly probable to virtual certainty; the law did not always recognize this limitation, but with time the law has acknowledged that the conceivable spectrum of standards coalesced irresistibly into three.[79] For another example, the harmless-error doctrine frequently invokes one of three low possibilities of an error's effect on outcome.[80] More generally, the law's standards of decision invoke a coarsely gradated scale of likeli-

---

exposed to a lethal pathogen.'" Andrew Taylor, *Congressional Offices Receive Mailed Threats*, YAHOO! NEWS (Feb. 23, 2012, 1:47 AM), http://news.yahoo.com/congressional-offices-receive-mailed-threats-220538717.html.

77   *See* Richard Bellman & Magnus Giertz, *On the Analytic Formalism of the Theory of Fuzzy Sets*, 5 INFO. SCI. 149, 151–52 (1973) (showing the equivalence between fuzzy sets and fuzzy statements).

78   *See* Clermont, *supra* note 20, at 1116–34 (drawing examples from such areas as standard of proof, standard of review, harmless error, trial motions, police actions, and administrative law).

79   *See* J. P. McBaine, *Burden of Proof: Degrees of Belief*, 32 CALIF. L. REV. 242 (1944) (arguing that, for standards of proof, only three levels of strength exist; using "degrees of belief" in Bentham's sense of strength of belief); C.M.A. McCauliff, *Burdens of Proof: Degrees of Belief, Quanta of Evidence, or Constitutional Guarantees?*, 35 VAND. L. REV. 1293 (1982).

80   *See* Clermont, *supra* note 20, at 1121–23.

hood stretching across the broader spectrum from the slightest possibility up to virtual certainty.

The reason for this coarse gradation, I argued, lay in the cognitive psychology literature. Cognitive limitations leave humans able only weakly to judge likelihood on any sort of scale. Studies of humans' weak absolute judgment,[81] restricted short-term memory, and use of biased heuristics all supported the limited capability of humankind. Those studies suggested that a step-like scale of intervals accords with how humans naturally process such information: judged likelihood customarily finds expression in terms of a very small set of broad verbal categories. Today, in all the probability and logic theories, there seems to be an emerging sense of the need to confront the limited precision of humans in gradating their beliefs.[82] It might therefore be more psychologically feasible for the law to ask fact-finders for an approximate degree of truth than for their precise view of probability.[83] Perhaps the law has already optimized by intuitively conforming to the coarsely gradated scale of likelihood already in people's customary use:

> The law usually does, realistically can, and optimally should recognize only seven categories of uncertainty in its standards of decision: (1) slightest possibility, (2) reasonable possibility, (3) substantial possibility, (4) equipoise, (5) probability, (6) high probability, and (7) almost certainty. First, this essay's description of seemingly diverse legal doctrines demonstrated that standards of decision tend to fall, often in groups of three, into the seven customary categories. Second, a review of cognitive psychology revealed humans to be "boundedly rational." Third, combining the observation with the science suggested that the systematic structure of the standards reflects the law's wise reconciliation with those cognitive limitations.[84]

Now, I espouse expressing that conclusion in the terms of fuzzy logic. I propose viewing the seven gradations as degrees of truth in

---

81 Absolute judgment involves reference to a remembered scale. Although not entirely distinct, relative judgment concerns the considerably greater capacity of people to distinguish between two or more different stimuli that they can compare directly. *See* WILLIAM N. DEMBER & JOEL S. WARM, PSYCHOLOGY OF PERCEPTION 113, 116–17 (2d ed. 1979).

82 *See* Terrence L. Fine, *[The Axioms of Subjective Probability]: Comment*, 1 STAT. SCI. 352, 353 (1986).

83 *See* Clermont, *supra* note 20, at 1139–44 (recounting empirical evidence); Yanlong Sun et al., *Probabilistic Judgment on a Coarser Scale*, 9 COGNITIVE SYS. RES. 161 (2008) (recounting more recent, consistent results); *supra* text accompanying note 50.

84 Clermont, *supra* note 20, at 1156.

this way (although I would redraw the separate gradations to be equally sized and shaped):
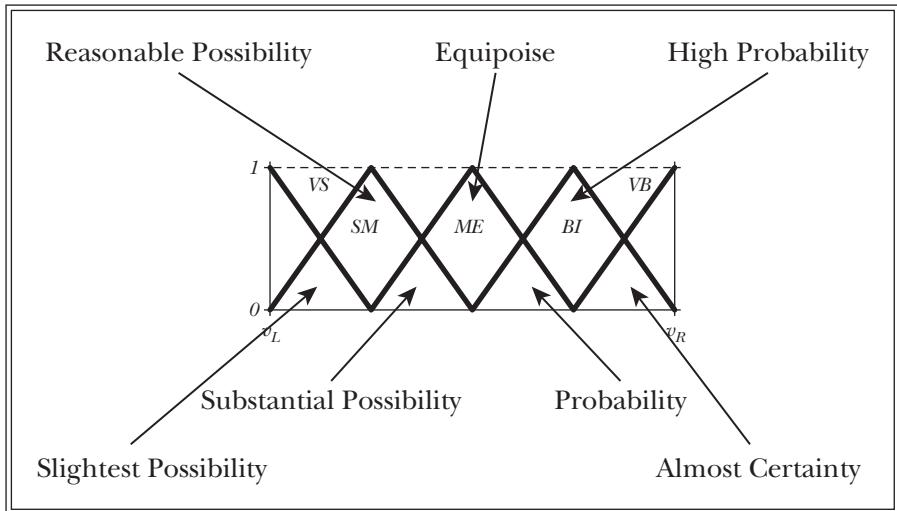
Thus, fuzzy logic accommodates the cognitive limitations of humans. Fuzzy logic offers a rich approach to elaborating the law's gradated scale of likelihood. Its real value, however, is that it captures the epistemic function in law better than probability. We want to know the belief in truth of a factual element, not the chance that the element will turn out to be 100% true.

II.   CONJOINING ASSESSMENTS

This Part will explain that probability's product rule for conjoined events does not apply in fuzzy logic. Then, this Part will show how the law relies on fuzzy logic when it applies the standard of proof to each element of claims or defenses, without worrying about applying the standard to conjoined elements.

A.   Fuzzy Operators

The power of the fuzzy conceptualization becomes more obvious when one considers the combination rules of fuzzy logic. These can become quite complicated, but for our purposes those of special interest are the most basic fuzzy operators, or connectives.

One constructs any system of logic by stipulating a small but adequate number of logical operators, such as intersection (or conjunction or $\wedge$ or AND), union (or disjunction or $\vee$ or OR), and negation

(or ~ or ¬ or NOT). They are the premises that generate an internally sound and complete system.[85]

1.   Maximum and Minimum

a.   Classical Logic's Operators

This bivalent system, which recognizes only the two values of true and false, stipulates the following functions for intersection and union:

truth(x AND y) = 1 if both x and y are true, but 0 otherwise

truth(x OR y) = 1 if either x or y is true, but 0 otherwise

Another way to state these two functions is this:

truth(x AND y) = minimum(truth(x), truth(y))

truth(x OR y) = maximum(truth(x), truth(y))

A different format in which to stipulate an operator is by truth table. The one for negation indicates that the negative of 1 is 0, and vice versa:

~	
1	0
0	1

All things bivalently logical flow from these three stipulations.

b.   Fuzzy Logic's Operators

Those three operators for fuzzy logic are just the same, except that they must extend to give results for values between 0 and 1.[86]

---

85   *See* SIDER, *supra* note 6, at 25, 35–37, 67–80 (showing also that in going beyond two-valued logic, one needs to stipulate the implication operator as well).

86   *See* Brian R. Gaines, *Fuzzy and Probability Uncertainty Logics*, 38 INFO. & CONTROL 154 (1978) (showing that the operators for fuzzy logic and probability theory are the same until one adds the assumption of the excluded middle).

Thus, the AND and OR functions work this way for sets in fuzzy logic, when x and y can take any value from 0 to 1:[87]

truth(x AND y) = minimum(truth(x), truth(y))

truth(x OR y) = maximum(truth(x), truth(y))

So, let X be the universe, and let A be one fuzzy set and B be another fuzzy set in the universe. The two sets might be independent, in the sense that the degree of membership in one set has no effect on the degree of membership in the other set, but they need not be. Designate the membership of element x in A as truth(x), and the membership of element y in B as truth(y). Then, the truth of the conjunction of x and y equals the smaller of the truth of x and the truth of y.

For an example involving a common element, let X be the universe of men, and let A be the set of tall men and B be the assumedly independent set of smart men. So, if Tom is a .30 member of A and a .40 member of B, then Tom is a .30 member of the set of tall and smart men. The intersecting set becomes smaller, but Tom's degree of membership in it does not decrease below the lower of his tallness and smartness levels. In other words, the truth value for the intersection would be the minimum value of the two memberships in A and B.

The following diagram may help to visualize this so-called MIN operator by indicating the shaded intersection of the two sets, where μ gives the degree of membership of an element in the fuzzy set. Along the x-axis, for any z that falls in the intersection, the degree of membership therein will be the degree of membership in A or B, whichever has the lower membership line at that point z:[88]

---

87   For elaborations of fuzzy intersection and union, see Radim Bělohlávek et al., *On the Capability of Fuzzy Set Theory to Represent Concepts*, 31 INT'L J. GEN. SYS. 569, 575 (2002); Ronald R. Yager, *Connectives and Quantifiers in Fuzzy Sets*, 40 FUZZY SETS & SYS. 39 (1991).

88   The figure comes from Ross & Parkinson, *supra* note 17, at 33; *see also id.* at 34–36 (extending the operator from an element's membership in multiple fuzzy sets to the relationship of different elements' memberships in different fuzzy sets).
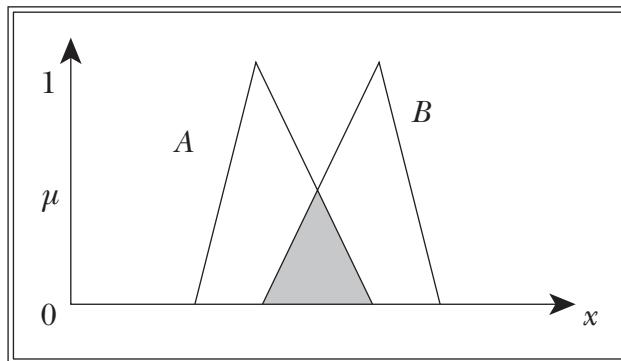
c. Justifying Logical Operators

Now, one can generate a logic system from any adequate group of operators. It will be internally sound on a formal level, but it will not be useful unless the operators make sense in our world. What makes sense is a philosophical question. But philosophers have to punt on this question, saying that operators make sense if they proceed from "genuine logical truths" and if their formal logical consequences constitute "genuine logical consequences."[89]

There are several signs that fuzzy logic makes sense. To begin, fuzzy logic does not produce nonsensical results. For example:

> The simplest and most fundamental qualitative law of probability is the extension rule: If the extension of *A* includes the extension of *B* (i.e., $A \supset B$), then $P(A) \geq P(B)$. Because the set of possibilities associated with a conjunction *A&B* is included in the set of possibilities associated with *B*, the same principle can also be expressed by the conjunction rule $P(A\&B) \leq P(B)$: A conjunction cannot be more probable than one of its constituents. This rule holds regardless of whether *A* and *B* are independent . . . . Furthermore, it applies not only to the standard probability calculus, but also to nonstandard models . . . .[90]

The MIN rule in fuzzy logic conforms to the extension rule by setting the conjoined probability of elements, whether or not independent, as equal to the least likely element.

More than that, the MIN rule affirmatively makes sense as the way for conjoining multivalent values. Tom really appears to be a .30

---

89 *See* SIDER, *supra* note 6, at 1–2, 6–11.

90 Amos Tversky & Daniel Kahneman, *Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment, in* HEURISTICS AND BIASES 19, 20 (Thomas Gilovich et al. eds., 2002) (emphasis omitted).

member of the set of tall and smart men. It is therefore the way to combine truth degrees more generally.[91]

Furthermore, fuzzy logic is not wildly different from classical logic. It does not require a radical overhaul of worldview. The choice posed is between (1) fuzzy logic and (2) bivalent logic with its probability overlay. In essence, fuzzy logic says only that we should account for the undeniable imprecision of the world by altering the system's operators, rather than by some awkward afterthought squeezed into the probability calculus.

At bottom, though, fuzzy logicians are arguing that their logic is different because it makes more sense than classical logic. "There are many reasons to get interested in nonclassical logic, but one exciting one is the belief that classical logic is *wrong*—that it provides an inadequate model of (genuine) logical truth and logical consequence."[92] The argument is that classical logic, by assuming the principle of biva-

---

91   The proof would go as follows. Reasoning backward from what is necessary for a system to make sense,

$$x \wedge x = x \qquad (1)$$
$$x \vee x = x \qquad (2)$$
$$x \wedge y \leq x \qquad (3)$$
$$x \vee y \geq x \qquad (4)$$

while associativity and distributivity need to prevail as well,

$$(x \wedge y) \wedge z = x \wedge (y \wedge z) \qquad (5)$$
$$x \vee (y \wedge z) = (x \vee y) \wedge (y \vee z) \qquad (6).$$

Then, using (2) and (3),

$$x \wedge (x \vee y) = (x \vee x) \wedge (x \vee y) \leq x$$

and, using (4) and (6),

$$x \vee (x \wedge y) = (x \vee x) \wedge (x \vee y) \geq x$$

and their having been shown to be equal, and both $\leq$ and $\geq x$,

$$x \wedge (x \vee y) = x \vee (x \wedge y) = x \qquad (7).$$

Now, designate y as the lesser or equal of the two truth values x and y. There should be a z such that $x \wedge z = y$, which allows the final conversions with the use of (7) and of (5) and (1), respectively:

$$x \vee y = x \vee (x \wedge z) = x = \mathrm{MAX}(x, y)$$
$$x \wedge y = x \wedge (x \wedge z) = (x \wedge x) \wedge z = x \wedge z = y = \mathrm{MIN}(x, y).$$

See Bellman & Giertz, *supra* note 77, at 151–55 (proving that the MIN and MAX operators "are not only natural, but under quite reasonable assumptions the only ones possible" for fuzzy sets); D. Dubois & H. Prade, *A Review of Fuzzy Set Aggregation Connectives*, 36 INFO. SCI. 85, 89–92 (1985) (showing that conjoined membership must of course be less than or equal to the minimum membership, but that accepting a value less than that minimum would produce nonsensical results).

92   SIDER, *supra* note 6, at 72.

lence, assumes one too many logical truths. It assumes a world where everything appears as on the left in this figure:[93]
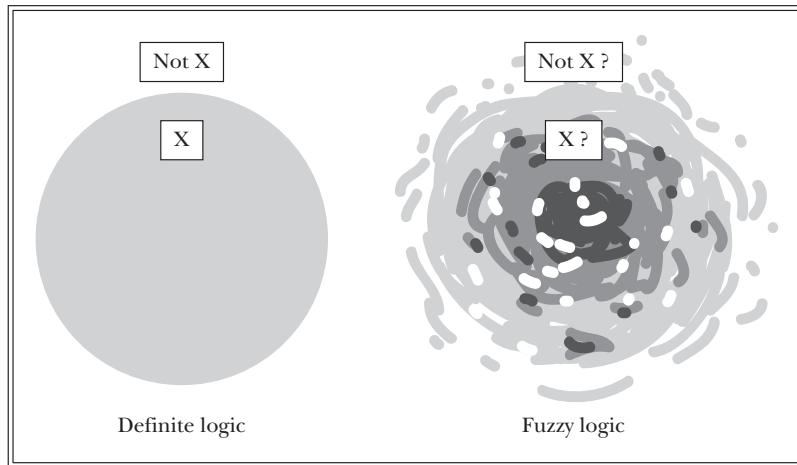
Like Euclidean geometry and Newtonian physics, classical logic is very useful, but an oversimplification.

2.   Product Rule Contrasted

a.   Applying Different Rules

The reader should nevertheless be sensing that something odd is afoot. Indeed, this is where most readers will abandon ship. After all, the probability operation for AND is multiplication of the probabilities of independent events.[94] But fuzzy logic tells us to apply the MIN operator even for independent events.[95]

Think of a room with ten men, each five feet-six inches tall. We might think of each as .30 tall. What would we term the tallness of the ten men as a group? It would still be .30 by the MIN operator. It would not be $.30^{10}$, a very tiny number yielded by the product rule to reflect the remote chance of them all turning out to be truly tall.

---

93   The figure comes from 3 Avi Sion, Logical and Spiritual Reflections ch. 4 (2008), *available at* http://www.thelogician.net/6_reflect/6_Book_3/6c_chapter_04.htm.

94   For interdependent events, the probability operation for conjunction is P(A) multiplied by P(B|A). Meanwhile, fuzzy logic tells us to apply the MIN operator, which is so much easier to comprehend and apply.

95   Analogously, De Morgan's rule provides that the product rule works for the OR operator on probabilities too: the disjunction of two independent statements equals the negation of the conjunction of the negations of those statements. Meanwhile, fuzzy logic tells us to apply the MAX operator.

Now if the room has ten men, with three "short" men five feet-six inches or below and three "dumb" men, then one man picked at random has a .09 chance of being both short and dumb, assuming independence. So, here the product rule applies.

That is, the probability operator is not inconsistent with the fuzzy operator. The two just contemplate different contexts. Indeed, the two are fundamentally consistent, because traditional probability is a special case of fuzzy logic's more general theory of uncertainty. So, it is not that one of these theories of uncertainty is correct and the other is wrong. It is that one theory can include the other.

For random uncertainty in a bivalent world, the probability operator will give the right answer, but so would the MIN rule. First, if the world were crisp, and x and y were known to be either true=1 or false=0, then their conjunction would be either 1 if both were true or 0 if not. In this narrow setting, the probability and fuzzy operators are equivalent. That is, the product rule would be no different from the MIN operator: truth$(x) \bullet$truth(y)=minimum(truth(x), truth(y)). Second, the chance of complete conjunction of currently unknown variables—that x and y will both turn out independently to be 1, or completely true—will be the product of their individual probabilities in either logic system. The product will make most sense in connection with frequentist probabilities. Still, the uncertainty could concern unique events, because one can speak of the subjective probability of x and y turning out to be either true=1 or false=0.

Remember that there are multiple kinds of uncertainty, including the indeterminacy resulting from scarce information or conflicting evidence and also the uncertainty characterized as either vagueness or randomness. If one tries to deal with the variedly uncertain real world, the more inclusive approach to conjunction becomes appropriate. In a fuzzy world, the product rule retreats to a specialized role, applying only when the independent values of x and y happen to be randomly uncertain without being vague. The product of probabilities gives the chance of things, which can take only a value of 1 or 0, coming up as 1 in two independent trials under conditions of random uncertainty. The intersection of degrees of truth is telling you how much you believe two statements put together. The latter is more general.

In sum, the product rule is not a feature only of classical logic. Both under classical logic with a probability overlay and under fuzzy logic, the MIN rule will reduce to the product rule—if one assumes bivalence and then adds an assumption of random independence. But the product rule will prevail under either system only if the elements under consideration are always ascertainable to be completely

true or completely false. Thus, the question becomes whether one should so assume bivalence.

b. Choosing Between Rules

Because both the product rule and the MIN operator can give correct, but sometimes different, answers, they must be giving answers to different questions or, rather, questions resting on different assumptions. The product of probabilities is answering a question different from what the intersection of degrees of truth is answering. The nature of the desired answer will determine the correct question to ask and, hence, whether the product rule or the MIN operator is appropriate to apply.

First, as a thought experiment, ponder which is the correct question to ask when one wants to know if Tom is tall and smart. Begin with the two membership statements given above—Tom is a .30 member of A and a .40 member of B. Those numbers mean something like "Tom is not so tall" and "Tom is not so smart."

The fuzzy combination would yield: "Because Tom is not so tall and Tom is not so smart, Tom is not such a tall, smart man." The MIN operator yields a .30 belief in that intersection. The traditionally probabilistic calculation, however, would yield: "Because Tom is not so tall and Tom is not so smart, Tom is likely a short, dumb man." The chance of a tall and smart Tom according to the product rule is .12, not .30, so that the product is lower than either truth(x) or truth(y).

This calculation by the product rule would be appropriate for certain kinds of decisions (and bets), but seems inappropriate for determining one's belief in Tom's membership in the set of tall and smart men. Multiplication of probabilities gives the chance that Tom is both completely tall and completely smart, while what we want to know is the degree to which he is both tall and smart. The inappropriateness becomes much more obvious as one combines more and more elements in the calculation, such as tall, smart, rich, and bald men. The product calculation will approach .00, even if some of the values are very high. The fuzzy combination, however, will go no lower than the minimum truth value. In other words, a fuzzy intersection of very true statements is very true, not almost completely untrue.

Second, one might try to classify a thing as a chair and as a red object. If the thing has some of the characteristics of a chair[96] and some pinkish hue, one would give it, perhaps, a .6 membership in the

---

96   *See* H. G. WELLS, A MODERN UTOPIA 381 (1905) ("I would undertake to defeat any definition of chair or chairishness that you gave me.").

set of chairs and a .5 membership in the red set. Now, if one had to give it a membership in the class of red chairs, one would say .5 for this reddish chair-like thing. One would not apply the product rule to say .3.

When would one apply the product rule for probabilities? One would do so when things are completely chairs or not and red or not, and you cannot see the thing, but you have an idea of the likelihood of chairishness and redness. To compute the chances of what bivalent values one will see when the thing is uncovered, and the thing becomes clearly a chair or not and red or not, one would use the product rule.

Many sorts of legal situations call for the product rule. In manipulating and evaluating statistical evidence, the fact-finder would often use it.[97] In calculating the odds of future events, as in computing expected costs on a motion for a preliminary injunction, the product rule would be appropriate.[98] There is a proper realm for the

---

[97]   A colleague gave this illustration:

Suppose that the Black Death strikes some town in England in 1349. Let's suppose that by the end of the year it kills 500 of the 1000 people then living in the town. A historian today is interested in figuring out whether ten particular people who lived in the town at the beginning of 1349 were killed by the Black Death later that year. The historian searches through the cemeteries, through church records and through other materials but comes up empty. There is simply no specific credible evidence about how any of these ten died. The historian can't even figure out how old each of them was at the time and thus adjust the odds based on different survival rates for different ages. Accordingly, his best guess is that for each of the townspeople, there is 50% probability that he or she died from the Black Death. Now the historian wants to know what are the odds that all ten died from the Black Death. The product rule says it's 1/1024 (unless there's some reason to think these are connected events, like they shared a household, so let's assume no info is known about such things). Fuzzy logic says it's one in two, which seems very obviously wrong. Indeed, assuming again that we know nothing further about any of the inhabitants of the town, fuzzy logic would tell us that the odds that *everyone* in the town died from the Black Death are one in two, but we know—because we assumed it to begin the inquiry—that the odds that everyone in the town died of the Black Death are zero. Only half of the inhabitants died of the Black Death. This seems to me a proof by contradiction of the applicability of the product rule to this sort of case.

E-mail from Michael Dorf to author (June 3, 2012, 22:04 EST).

[98]   *See* John Leubsdorf, *The Standard for Preliminary Injunctions*, 91 HARV. L. REV. 525, 542 (1978) ("The court, in theory, should assess the probable irreparable loss of rights an injunction would cause by multiplying the probability that the defendant will prevail by the amount of the irreparable loss that the defendant would suffer if enjoined from exercising what turns out to be his legal right. It should then make a similar calculation of the probable irreparable loss of rights to the plaintiff from deny-

product rule, just as there is one for the straightforward application of the MIN rule. The question before us is whether a significant share of legal applications of the standard of proof falls into the latter realm.

Third, picture a column of one hundred coins, thirty of them heads randomly placed, and another column of one hundred coins, forty of them heads randomly placed. Then only about twelve paired rows will have two heads. Or picture a column of one hundred people, thirty of them tall people randomly placed, selected from a universe where people are either completely tall or completely short; and picture another column of one hundred more people, forty of them smart people randomly placed, selected from a universe where people are either completely smart or completely dumb. Then only about twelve paired rows will be persons tall and smart, respectively. Now, picture instead a column of varying beliefs in the tallness of one hundred people selected from a universe where people have the tallness trait distributed naturally, aligned from tall down to short, and another column of one hundred beliefs about persons, aligned from smart down to dumb. The beliefs concerning the thirtieth pair from the bottom will be not so tall and not so smart, respectively, while the twelfth pair from the bottom will be a diminutive dim couple.

c. The Law's Choice

Traditional probability and degrees of truth do therefore differ. They behave differently in the conjunction setting. Put simply, the product rule gives the random chance of the simultaneous and independent occurrence of multiple crisp elements, while the MIN operator measures the intersection of sets. Once lawmakers have in mind the difference between the product rule and the MIN operator, they have to decide which the law should apply.

Imagine the law is trying to determine if Tom himself was at fault, that is, whether the perpetrator was Tom and whether the perpetrator was at fault. A number of features of this compound question indicate that the better approach for law is fuzzy logic. First, the two parts of the question are epistemically very different, one being a factual event and the other a legal construct; the law needs commensurable measures to combine them. Second, as already argued, the law should not be worrying too much about the chance of a truth value of 1 turning up; instead it should ascertain whether the element has a sufficient truth value. Third, in establishing past truths, the law should be even less concerned with the chance of a 1 repetitively turning up; applying

---

ing the injunction. Whichever course promises the smaller probable loss should be adopted." (footnote omitted)).

the product rule to subjective probabilities for legal fact-finding actually seems illogical.[99]

That third point is indeed determinative. One can similarly make the point by distinguishing between ex ante probabilities, used to make predictions of what you will eventually know, and ex post probabilities, used to decide what actually happened even though you will never know for sure. If you are placing a bet predicting whether two randomly uncertain events will independently happen together, then multiply their probabilities. But if you are looking back to the past, then you need a different operator. You are no longer trying to figure the odds of two things being sure, but rather how sure you are that one thing happened while you remain somewhat sure that the other happened. "The ex post probability for complete instantiation of the causal law is equal to the lowest ex post probability for instantiation of any constituent element."[100]

In other words, we want to know if it is a reddish chair, not what the chances are that it is 100% a chair and also 100% red. We want to know if Tom's fault is sufficiently true, not the chances of somehow discovering both the perpetrator certainly to be Tom and the perpetrator to be completely at fault. Here is another way to see this. If Tom is 60% likely the perpetrator, and the perpetrator is 70% at fault, the 60% figure means that it is 60% likely that Tom is surely the person who was 70% at fault. We thus have a 60% chance of Tom's being legally at fault, just as the MIN rule would say.

While the MIN rule seems the obvious choice if identity is a matter of occurrence uncertainty and fault is a matter of imprecise vagueness, I think it still should apply even if both fact-finding percentages measure only random uncertainty. Having different operators for different kinds of elements, leading to some weird hybrid calculation unknown to current law, would be more than awkward. But I am arguing that the MIN rule is the right approach, not just a convenient one. A 60% chance of the weakest link represents the chance that all the other elements are more likely than not to exist. Because a 70% chance of fault is good enough for liability, we should not further account for that chance of finding complete fault. To multiply the chances, getting 42%, would be double counting, as it represents the chances of fully establishing both identity and fault. The chances of

---

99   *See* Didier Dubois & Henri Prade, *A Set–Theoretic View of Belief Functions: Logical Operations and Approximations by Fuzzy Sets, in* CLASSIC WORKS OF THE DEMPSTER–SHAFER THEORY OF BELIEF FUNCTIONS, *supra* note 55, at 375, 403 (rejecting the application of "arguments deriving from the study of statistical experiments").

100   Richard W. Wright, *Proving Facts: Belief Versus Probability, in* EUROPEAN TORT LAW 2008, at 79, 93 (Helmut Koziol & Barbara C. Steininger eds., 2009).

proving completely each of multiple elements simultaneously would be an odd inquiry when the law does not demand complete proof of any. Because establishing every element to 100% is not what the law calls for, the chances of doing so are irrelevant. The relevant inquiry comprises how likely the weakest element is, given that all the other elements would simultaneously be stronger.

Provability versus probability—Reactions of colleagues have convinced me that elaboration of this assertion, even in multiple ways, is necessary. So, up to this point, I have established that the MIN and product rules are both valid operators, but they govern in different realms. Which, then, should govern in applying the standard of proof?

In explaining the law's choice, let me begin with the contrary intuitive yearning to apply the product rule. If element *A* is 60% likely and element *B* is 70% likely, if both *A* and *B* either occurred or did not, and if the law wants to allow recovery only if *A* and *B* both occurred, then it does seem that the plaintiff should lose on this 42% showing.

My initial counterargument is that this result is tough on plaintiffs. Multiple elements would stack the deck against real-world plaintiffs, who must live with the imperfections of available evidence. Imagine some other plaintiff having proven four elements each to 70%. That plaintiff has done a really good job in presenting a strong case. The plaintiff has well established each element before passing to the next one. The plaintiff has done exactly what we should demand. Yet this plaintiff would lose with a miserable 24% showing under the product rule. What happened? How did a strong case become a sure loser? Regardless of how, plaintiffs apparently would lose strong cases they really should win. Moreover, defendants at fault would not be receiving a corrective message. These errors would detrimentally affect economic efficiency.

Perhaps the law should not interest itself in the probability of the elements all turning out to equal 1, if only the veil on perfect knowledge were lifted. If every event and thought were somehow videotaped, then we would be partially living in an ascertainably bivalent world, and the law's approach to standards of proof might have to change. But I think the law should not imagine the real world to be a videotaped one. Adopting a false assumption simply in order to make a familiar math tool available is usually indefensible. The law should not ask the odds of the elements bivalently and conjoinedly existing on the videotape. The law instead should ask how well the burdened party has proven its case. The fact-finder needs to operate on the

basis of its resulting internal beliefs about the world, rather than pretending that external knowledge is attainable.

Provability, not probability, is the law's concern. Forming a belief as to what happened, rather than a prediction about a veil-lifting that will never happen, is the aim. The law wants to know whether the state of our knowledge based on proof justifies recovery. Fuzzy logic vaguely deals with the "probably provable," while traditional probability crisply deals with the "provably probable."[101]

Expression of this provability comes in terms of membership, to a degree, in the set of true statements, with the degree measured as a truth value. Provability of one element does not detract from the provability of another. An easily provable *A* and an easily provable *B* mean that it will be easy to prove *A* and *B*. The intersection of sets represents the interaction of elements, and the MIN rule governs the intersection of sets. Consequently, if the plaintiff proves element *A* to 60% and *B* to 70%, then the provability that the case is in the set of (*A* AND *B*) is 60%.

The plaintiff has shown that the conjoined claim is 60% provable and the defense 40% provable. That is, the belief in the claim is stronger than its negation (the belief that one or the other element or both elements failed). To minimize errors, the law should decide in conformity with the stronger belief. If the law were to deny liability in these circumstances because of some attraction to bivalent probability theory, more often than not the law would be wrong. Giving the plaintiff a recovery and the defendant a loss thus is economically efficient. Accordingly, the law should and does instruct the use of the mathematically sound way to combine beliefs, here the MIN rule.

This key distinction between probability and provability was at the heart of Oxford philosopher L. Jonathan Cohen's almost impenetrably brilliant book entitled *The Probable and the Provable.*[102] He argued that the task of the law court is to decide, by use of inductive reasoning, what is provable. Importing traditional probability into the pro-

---

101   *Cf.* Schum, *supra* note 22, at 243 (discussing belief functions, and crediting Judea Pearl, *Bayesian and Belief-Functions Formalisms for Evidential Reasoning: A Conceptual Analysis, in* Readings in Uncertain Reasoning 540, 571 (Glenn Shafer & Judea Pearl eds., 1990), for this phrasing).

102   L. Jonathan Cohen, The Probable and the Provable (1977), *reviewed by* David A. Schum, 77 Mich. L. Rev. 446 (1979), *and* Carl G. Wagner, 1979 Duke L.J. 1071; *see also* Bertrand Russell, Human Knowledge: Its Scope and Limits 359–61 (1948) (arguing comparably that his "degrees of credibility" do not follow the rules of traditional probability); Susan Haack, *The Embedded Epistemologist: Dispatches from the Legal Front*, 25 Ratio Juris 206, 217–18 (2012) (arguing comparably that her "degrees of warrant" do not follow the rules of traditional probability).

ject, such as the product rule, produces a whole series of anomalies. Instead, the conjunction rule for inductive reasoning is this: "The conjunction of two or more propositions . . . has the same inductive probability . . . as the least" likely conjunct.[103]

Thus, the respective realms of the MIN and product rules do not turn on the nature of the fact issue, but on the question the system wishes to pose. Which image fits the fact-finding endeavor: the betting table or set theory? I think that standards of proof are looking for provability based on set theory. They therefore take the same approach to facts involving occurrence uncertainty as they do on to facts involving vagueness.

Monty Hall's contribution—Why do smart people so resist accepting that provability differs from probability? "When the plaintiff proves one element to 60% and another to 70%, their conjunction is 42%—and I am sticking to it!" This reaction irresistibly brings to mind the usual reaction to the celebrated Monty Hall problem. "It is customary for books about probability to try to persuade otherwise intelligent people that they are lousy when it comes to reasoning about uncertainty. . . . In presenting the Monty Hall problem to students I have found the common reactions to follow the well-known five stages of grief."[104] There is denial, anger, bargaining, depression, and then acceptance.

Consider the related "sibling gender problem."[105] A few years back you saw Tom walking down the street with his son. Your companion said that she remembers he has two children. What are the chances that the other child is a boy? The answer is one-third, because the equally probable possibilities are BB, BG, and GB. But if your companion had said that the elder child was a boy, the answer would be one-half! The additional information, seemingly irrelevant, provides ordering that affects the odds.

---

103   COHEN, *supra* note 102, at 266.

104   JASON ROSENHOUSE, THE MONTY HALL PROBLEM 5 (2009). The literature here is immense. One nifty entry was a report on how overlooking the additional-information effect on probabilities had invalidated decades of research on cognitive dissonance. *See* John Tierney, *And Behind Door No. 1, a Fatal Flaw*, N.Y. TIMES, Apr. 8, 2008, at F1, *available at* http://www.nytimes.com/2008/04/08/science/08tier.html ("Even some of the smartest mathematicians initially come up with the wrong answer to the Monty Hall Problem. Perhaps the best way to understand it is to play the game yourself."). The web-version of the article links to a site, http://www.nytimes.com/2008/04/08/science/08monty.html#, that allows you to play the game repetitively and so build to the right strategy.

105   *See* ROSENHOUSE, *supra* note 104, at 26, 138–41, 147–48.

After you have progressed through a couple of the stages of grief toward acceptance of that result,[106] consider that traditional probability is generating all those emotions. When a problem calls for rejecting bivalence, you should expect that sometimes the answer will be similarly nonobvious. For example, reconsider a plaintiff trying to prove the identity of the perpetrator being Tom and also to prove the perpetrator being at fault. True, if the randomized odds are 60% and 70%, the odds of Tom being at fault are 42%. The product rule gives that result. But if the plaintiff has proved fault to 70%, the odds on the remaining question of Tom being the perpetrator are 60%. The MIN rule sets the likelihood of the conjunction at 60%.

Using the setting of the more familiar "Bertrand box paradox"[107] for elaboration of the shift to multivalence, imagine that there is an identity box and a fault box, each containing a ball that is either black for liability or white for non-liability. The two balls came, respectively, from an urn with 600 of 1000 balls being black and from another urn with 700 of 1000 being black. The odds of the two balls both being black are 42%. But if you uncover or otherwise decide that the fault box has a black ball, the odds of the identity ball being black are 60%.

What is going on? The adherents of 42% are assuming that the pairings are randomized. But in inductively proving a case—by establishing two truth values or fuzzy provabilities—the plaintiff was ordering the information.[108] The plaintiff thereby removed the randomization feedback loop between the boxes, a randomization of information essential to the 42% calculation.

Under its standard of proof, the law has decided to act on the basis of a partially proved case, not on the basis of the probability of a fully proved case. Fault proven to 70% will never convert to 1 or 0.

---

106   If you resist the result, and persist with two-out-of-three odds that the younger child is a girl, I propose the following gamble to you. You will bet on whether a hidden flipped coin is heads. But before placing the bets, I flip another coin, and it comes up tails. You then should believe there is a two-thirds chance that the hidden coin is heads, and so should offer me better than even money.

The effect of additional information emerges from this sequence: Flip two coins. What are the odds that they will both be heads? One-in-four. If you know one of them was heads, what are the odds they both were heads? One-in-three, because knowing the result of one flip tells us something about the other. If instead you know the first flip was heads, what are the odds they both were heads? One-in-two.

107   *See* Rosenhouse, *supra* note 104, at 14–16. Three boxes respectively contain two black balls, two white balls, and one black ball and one white ball. You pick a ball from one box, and it is black. What are the odds that it came from the mixed box? The answer is one-in-three.

108   Recall my illustration of the two ordered columns of tall and smart people, *supra* text accompanying note 98.

That means that 30% of the results are not findings of nonfault, but erroneous failures to find fault. Fault having satisfied the standard of proof, the 30% of pairings with nonfault then become errors. To minimize errors, the fact-finder should consider only the 70% of pairings of identity with established fault. Because 60% of those pairings will result in liability and 40% not, deciding in line with the 60% showing will minimize errors and optimize efficiency.[109]

Role of assumptions—There lies the key to the paradox. If one assumes bivalence, then one must convert fault to 1 or 0 before proceeding. If one instead recognizes multivalence, one can proceed with fault standing as a partial truth. The mathematically sound way to conjoin partial truths is the MIN rule. Therefore, recognition that the plaintiff can prove any element only to a degree produces an element-by-element approach.

Instinctive resistance to the MIN rule derives from residual yearning to apply a multiplicative rule of traditional probability to a problem it cannot handle, the problem of fuzzy provability. It can handle only randomly uncertain estimates of independent events in a binary world, because it is built on the logical assumption of bivalence.

---

109   Another colleague, after putting aside problems of market share and statistical evidence, challenges me to

> assume the plaintiff's decedent took a drug either from manufacturer D1 or from manufacturer D2 (the drugs are identical). Assume 60% probability of D1 and 40% of D2. Assume further that it is 60% likely that the drug (from whichever manufacturer) actually caused the death. So 4 possibilities: D1's drug caused the death (36%); D2's drug caused the death (24%); neither D1 nor D2 caused the death (40%). Why should P collect against D1?

E-mail from George Alan Hay to author (June 7, 2012, 11:15 EST) (names of parties altered).

Professor Hay is conducting a thought experiment, in which we pull off the veil to reveal a bivalent scheme and then randomly distribute the cause over the identity results. The drug as cause of death sometimes becomes a 1, but in forty out of one hundred cases it will become 0. The zeros fall randomly, instead of dropping out.

The thought experiment has no relevance to what the law or economic theory should do based on the actual proof, however. The thought experiment changes the problem, changing it in ways that affect what law and economics would do with respect to D1's liability. In the case against D1, liability when the cause will randomly be either 1 or 0 is a different question from liability when the plaintiff has proved cause to 60%.

So, I am indeed saying that P has a 60% provable case against D1, and should win just as the law says. P enjoys a truth value of 60% on the proposition that D1 made the drug, and another of 60% that the drug caused the death. This means that the proposition that D1's drug caused the death is 60% a member of the set of true statements. In other words, given the current state of our knowledge, P has 60% of full proof against D1. To decide against P would be to favor a defendant with 40% of a defense.

When an assumption no longer prevails, one cannot apply the rules built on the assumption. We tend to forget that mathematical constructs operate only within their assumed system, and that the probability calculus assumes all events will take a value of either 1 or 0. Multivalence calls for new math. We must move up to MIN.

There is more at work in obscuring the picture. Even if one acknowledges that the multiplicative rule should apply only in an abstract world of bivalence, one will not intuitively sense the subtle shift in a problem's setting from a bivalent assumption to a multivalent reality in which the middle is no longer excluded. The shift can be almost imperceptible. But when the problem is finding facts of which one will never be sure, the picture must be painted in multivalent grays.

The bottom line is this: *as an artifact of bivalence, the product rule does not apply to subjective probabilities for fact-finding.* I therefore submit that degrees of truth behave more appropriately than classical logic and probability theory for the purposes of the standard of proof.[110] But, again, my central question is which representation the law employs. Here, as I have already said, and as I shall show in the upcoming resolution of the conjunction paradox, I am confident that it is degrees of truth across the board.

3. Negation Operator

Accepting the usefulness of fuzzy logic prompts interest in other fuzzy operators. Another basic one is negation. Here truth$(\text{not}\,x)$ =

---

110   I rest on the positive arguments for the appropriateness of fuzzy logic. Additionally, there are practical arguments against the product rule. It might be cognitively challenging to apply. *See* Porat & Posner, *supra* note 70, at 47–48. Or the effect of an element-by-element approach might offset the inefficiencies of other legal rules. *See* Alex Stein, *Of Two Wrongs That Make a Right: Two Paradoxes of the Evidence Law and Their Combined Economic Justification*, 79 Tex. L. Rev. 1199 (2001). I do not need to rely on these practical arguments.

Some such arguments, however, are just wrong. For example, some argue that for jury decision-making the necessity of convincing multiple fact-finders means, by virtue of the Condorcet theorem and the supermajority requirement, that the plaintiff's task is way too demanding; accordingly, to ameliorate the difficulty of proof, the system does not impose the additional demand of a product rule and instead proceeds element-by-element. *See* Saul Levmore, *Conjunction and Aggregation*, 99 Mich. L. Rev. 723, 734–45 (2001). This position rests on several errors. *See* Ronald J. Allen & Sarah A. Jehl, *Burdens of Persuasion in Civil Cases: Algorithms v. Explanations*, 2003 Mich. St. L. Rev. 893, 904–19; Paul H. Edelman, *On Legal Interpretations of the Condorcet Jury Theorem*, 31 J. Legal Stud. 327, 343–48 (2002). To me, the most obvious error lies in ignoring that a decision for the defendant also requires the agreement of the multiple fact-finders.

(1 - truth(x)), just as in classical logic. The negation is the complement.[111]

However, the whole fuzzy set A and its complement do not necessarily add to unity, because fuzzy logic does not obey the law of the excluded middle. The following figure demonstrates this fact, with the left graph representing the fuzzy set A by the solid line and its complement by the dotted line, and the right graph's dark upper line representing the union of A and its complement by operation of the MAX function:[112]
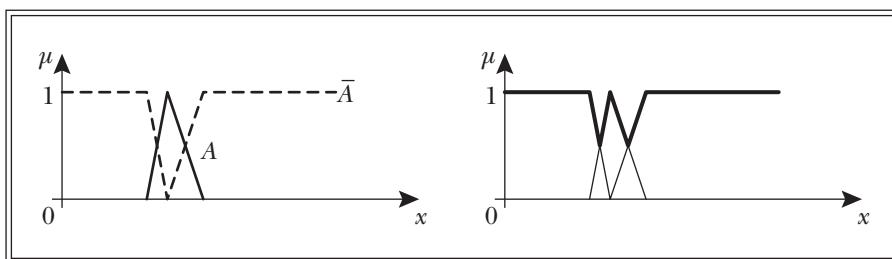
The serrations in the upper line in the right-hand graph show that A and its complement do not add to equal the universe X, reflecting that the law of the excluded middle does not hold. There will be an area where beliefs are neither in the set of belief nor in the set of disbelief, but instead are indeterminate.

B.   Legal Application: Conjunction Paradox

The payoff of the fuzzy logic approach emerges as one realizes how it affects the view of the proof process. Consider the best-known statement of the infamous conjunction paradox:

> We purport to decide civil cases according to a more-probable-than-not standard of proof. We would expect this standard to take into account the rule of conjunction, which states that the probability of two independent events occurring together is the product of the probability of each event occurring separately. The rule of conjunction dictates that in a case comprised of two independent elements the plaintiff must prove each element to a much greater degree than 50%: only then will the plaintiff have shown that the probability that the two elements occurred together exceeds 50%. Suppose, for example, that a plaintiff must prove

---

111   *But cf.* Bellman & Giertz, *supra* note 77, at 155–56 (showing that other negation operators are possible).

112   The figure comes from Timothy J. Ross, Fuzzy Logic with Engineering Applications 37 (3d ed. 2010).

both causation and fault and that these two elements are indepen-
dent. If the plaintiff shows that causation is 60% probable and fault
is 60% probable, then he apparently would have failed to satisfy the
civil standard of proof because the probability that the defendant
both acted negligently and caused injury is only 36%.

In our legal system, however, jurors do not consider whether it
is more probable than not that all elements occurred in conjunc-
tion. Judges instruct jurors to decide civil cases element by element,
with each element decided on a more-probable-than-not basis.
Once jurors have decided that an element is probable, they are to
consider the element established, repress any remaining doubts
about it, and proceed to consider the next element. If the plaintiff
proves each element by a preponderance of the evidence, the jury
will find in his favor. . . . Thus, jurors may find a defendant liable
even if it is highly unlikely that he acted negligently, that is, the
conjoined probability of the elements is much less than 50%. In
such cases, the verdict fails to reflect a probable account of what
happened and thus fails to minimize the cost of judicial errors. . . .

. . . .

. . . Although courts direct juries to consider and decide each
element seriatim, juries do not consider each item of evidence seria-
tim when deciding whether a given element is proved. The jury
must decide each element by looking at all of the evidence bearing
on proof of that element. Thus, although the jury does not assess
the conjunction of the elements of a case, it does decide each ele-
ment by assessing the conjunction of the evidence for it.[113]

---

113   Charles Nesson, *The Evidence or the Event? On Judicial Proof and the Acceptability of Verdicts*, 98 HARV. L. REV. 1357, 1385–88 (1985) (footnotes omitted). Professor Nesson saw the paradox as illustrating his broad thesis that the law's process of proof aims at generating acceptable *statements* about past events and thus at projecting behavioral norms to the public, rather than at reaching probable conclusions in a search for truth:

Application of the more-probable-than-not test to each element produces
the most acceptable conclusion as to that element. The conjunction of
these conclusions constitutes a story that is more probable than any other
story about the same elements. Suppose, for example, that the elements of a
story are *A* and *B,* and *A* (70%) is more probable than not-*A* (30%), and *B*
(60%) is more probable than not-*B* (40%). The conjunction (*A* & *B*) (42%)
may not be more probable than its negation (not-(*A* & *B*)) (58%). But the
conjunction (*A* & *B*) (42%) is more probable than any other version: (*A* &
(not-*B*)) (28%), ((not-*A*) & *B*) (18%), or ((not-*A*) & (not-*B*)) (12%). The
application of the more-probable-than-not standard of proof on an element-
by-element basis will produce the single most probable story.

Id. at 1389–90 (footnotes omitted). *See generally* J.S. COVINGTON, JR., THE STRUCTURE
OF LEGAL ARGUMENT AND PROOF 347–57 (2d ed. 2006) (discussing the misuses of
probability theory at trial).

The implications are profound but boggling. Allowing recovery on a 36% showing of causation *and* fault is not only unfair but inefficient. How embarrassing for the law!

For another boggle, ponder the apparent criticality of how exactly the ancients (and moderns) divided our causes of action and defenses into elements: the more subdivisions, the lower the conjunctive probability that would produce victory.[114] And yet:

> Anyone who has ever litigated a real case knows the exact opposite of the conjunction paradox is true: the more disputed elements the plaintiff has to prove, the *less* likely the plaintiff is to prevail. . . . [A]lthough it is possible that a particular plaintiff could obtain an unjust verdict in a case with several disputed elements, [there is an increased] probability that the jury will find at least one element to be less likely than not.[115]

Admittedly, the conjunction paradox turns out to be not such a serious problem in practice. Only one element might be in dispute, or the disputed elements might not be really independent. The judge might not clearly state, or the jury might not fully understand, the proper element-by-element approach to the standard of proof.

Or, because humans might tend to construct a story for the whole case rather than proceeding element-by-element, the fact-finder might end up applying the standard of proof to the conjoined elements. In fact, many psychologists agree that the fact-finder naturally constructs such stories, although perhaps not in a very systematic manner.[116] The broadly accepted *story model* of evidence processing holds that the fact-finder, over the trial process's course, constructs from the evidence the story that makes maximal sense; and the fact-finder then

---

114   *See* James A. Henderson, Jr. et al., *Optimal Issue Separation in Modern Products Liability Litigation*, 73 TEX. L. REV. 1653, 1655–59, 1667–75 (1995).

115   David A. Moran, *Jury Uncertainty, Elemental Independence and the Conjunction Paradox: A Response to Allen and Jehl*, 2003 MICH. ST. L. REV. 945, 946–47, 950.

116   *See generally* JEFFREY T. FREDERICK, THE PSYCHOLOGY OF THE AMERICAN JURY 296–99 (1987) (providing a brief overview of the story model of evidence processing); REID HASTIE ET AL., INSIDE THE JURY 22–23 (1983) (providing a brief summary of empirical studies supporting the story model); Paula L. Hannaford et al., *The Timing of Opinion Formation by Jurors in Civil Cases: An Empirical Examination*, 67 TENN. L. REV. 627, 629–32 (2000) (discussing "three predominant models of jury decision making"); Jill E. Huntley & Mark Costanzo, *Sexual Harassment Stories: Testing a Story–Mediated Model of Juror Decision–Making in Civil Litigation*, 27 LAW & HUM. BEHAV. 29, 29 (2003) (presenting research that "extends the story model to civil litigation and tests a story-mediated model against an unmediated model of jury decision-making"); Nancy Pennington & Reid Hastie, *The Story Model for Juror Decision Making, in* INSIDE THE JUROR 192 (Reid Hastie ed., 1993) (detailing the story model and summarizing empirical studies testing it).

chooses, among the available decisions, the one that fits best with the constructed story:

> Several authors have recently proposed a model for juror decisionmaking based on the concept of a story as an organizing and interpreting schema. The story model attempts to explain how jurors organize and interpret the vast amount of information they encounter at trial and apply the appropriate decision criteria. . . .
>
> . . . The jurors construct a story adequately describing what happened. At the conclusion of the trial, they construct the verdict categories based on the instructions given by the judge. The individual juror arrives at his decision by determining the best match between his story and the available verdict categories. The task of the jury in deliberations then becomes one of selecting a story from among those offered by the jurors and fitting it to the available verdict options.[117]

If the jurors construct a story (or stories[118]) for the whole case, or otherwise cognitively process the entirety while the trial progresses, and then the judge instructs on standard of proof, it might be that the jurors actually apply the standard to the whole claim or defense. It might also be that, being human, a judge when acting as fact-finder proceeds in essentially the same manner, testing whether the already conjoined elements are more likely than not.

Indeed, by providing obscure instructions only at the end of oral trials, the law seems determined to encourage overall consideration and to discourage applying the standard of proof element-by-element. Although the judge does instruct literally in element-by-element terms,[119] this may work only to encourage the jurors' detailed evalua-

---

117   FREDERICK, *supra* note 116, at 296–97 (citations omitted).

118   *Compare* Reid Hastie, *What's the Story? Explanations and Narratives in Civil Jury Decisions*, *in* CIVIL JURIES AND CIVIL JUSTICE 23, 31–32 (Brian H. Bornstein et al. eds., 2008) (expanding the theory to allow for a party's multiple stories), *with* Michael S. Pardo, *The Nature and Purpose of Evidence Theory*, 66 VAND. L. REV. (forthcoming 2013), *available at* http://ssrn.com/abstract=2060340 (discussing the theory's difficulties in handling multiple stories).

119   3 KEVIN F. O'MALLEY ET AL., FEDERAL JURY PRACTICE AND INSTRUCTIONS: CIVIL § 104.01 (6th ed. 2011):

> Plaintiff has the burden in a civil action, such as this, to prove every essential element of plaintiff's claim by a preponderance of the evidence. If plaintiff should fail to establish any essential element of plaintiff's claim by a preponderance of the evidence, you should find for defendant as to that claim.

See Allen & Jehl, *supra* note 110, at 897–904 (criticizing Dale A. Nance, Commentary, *A Comment on the Supposed Paradoxes of a Mathematical Interpretation of the Logic of Trials*, 66 B.U. L. REV. 947, 949–51 (1986) (finding this instruction ambiguous)).

tion of the evidence and to stress the requirement that any story must contain all of a series of elements—just as many evidence rules may work to brake any undesirable tendency of the fact-finder to rush toward creating a story.[120]

So, the conjunction paradox may not inflict great practical effects. Nonetheless, the big theoretical problem of the conjunction paradox will unavoidably pose at least some practical difficulties. The law sometimes enforces its element-by-element theory and thereby impedes the holistic practice. An obvious example would be when the judge requires a special verdict that asks the jury to find each element by a preponderance.[121] The conjunction paradox therefore remains troubling, and theorists twist themselves into pretzels trying to explain it away.

It would be troubling, however, only if theory really calls for the product rule. But theory does not. Instead, it invokes the MIN rule. The truth of the conjunction equals the minimum of the truths of the elements. If each element is more likely than not, then the truth of the conjunction is more likely than not. To use the above example, if the plaintiff shows that fault is .60 true and that causation is .60 true, then he has shown to .60 that the defendant *both* acted negligently and caused the injury.

Thus, there is no conjunction paradox. It implodes under the force of fuzzy logic. The MIN operator provides that belief in the conjunction will match the least likely element, which has already passed the standard of proof. The MAX operator meanwhile indicates that belief in the negative of the conjunction, that is, in the disjunction of each element's negation, will never reach equipoise. The story of liability will not only be the most believable story, but will be more believable than all the stories of non-liability combined.

Comfortingly, under the MIN rule, *applying the standard of proof element-by-element works out to be equivalent to applying it to the whole conjoined story.* So, if the fact-finder actually does follow the story model, that practice would not directly endanger the standard of proof. The apparent criticality of the number of elements melts away too. Because the MIN rule applies to each set of evidence to be conjoined,

---

120   *See* Bruce Ching, *Narrative Implications of Evidentiary Rules*, 29 QUINNIPIAC L. REV. 971 (2011) (discussing a narrative perspective in persuasion, and for evidentiary rules such as hearsay and party admissions); Doron Menashe & Mutal E. Shamash, *The Narrative Fallacy*, 3 INT'L COMMENT. ON EVIDENCE iss. 1, art. 3 (2006) (using narrative theory to criticize holistic evidence theory).

121   *But cf.* Elizabeth G. Thornburg, *The Power and the Process: Instructions and the Civil Jury*, 66 FORDHAM L. REV. 1837, 1857–63 (1998) (questioning whether a special verdict actually changes the jury's decision-making practice).

it does not matter where the law draws formal lines between elements, or whether the elements are independent or interdependent. Nor does it matter if I sloppily labeled identity as an "element" in my examples above.[122]

Moreover, the proof process within elements is not dissimilar to the proof process between elements. Within elements, the fact-finder uses intuitive techniques in a non-quantitative and approximate fashion. Between elements, and for separate facts within elements, the fact-finder uses the fuzzy operator for conjunction that works in a similar style.

The law does seem to know what it is doing, then. Whenever it phrases its instruction to require applying the standard of proof element-by-element, it is instructing to apply the MIN operator. But do actual fact-finders apply the MIN operator as they should and as the law tells them to do? We do not know. Some experimental evidence arguably suggests that the lay person tends to apply the product rule rather than the MIN operator.[123] Nevertheless, no sign exists that fact-finders in the legal system are using the product rule. After all, a concern that they were ignoring the product rule generated the unfounded fear of the conjunction paradox in the first place.

Theorists also claim there is a converse paradox, involving multiple theories. These observers lament that the law denies relief to a

---

122   *See* COHEN, *supra* note 102, at 267 ("So on the inductivist analysis, if the plaintiff gains each of his points on the balance of probability, he can be regarded as gaining his case as a whole on that balance . . . , without any constraint's being thereby imposed on the number of independent points in his case or on the level of probability at which each must be won.").

123   *See* Gregg C. Oden, *Integration of Fuzzy Logical Information*, 3 J. EXPERIMENTAL PSYCHOL.: HUM. PERCEPTION & PERFORMANCE 565, 568–572 (1977). His experiment involved having students judge the degree of truthfulness of statements like "a chair is furniture" and "a pelican is a bird," and asking them for the degree to which both statements together were true. The students seemed to use the product rule rather than the MIN rule. But it seems to me that the students could have interpreted these statements as verifiably being either completely true or completely false, thus making the product rule appropriate. Moreover, other experiments indicate that people do use fuzzy operators. *See* Rami Zwick, David V. Budescu & Thomas S. Wallsten, *An Empirical Study of the Interpretation of Linguistic Probabilities*, *in* FUZZY SETS IN PSYCHOLOGY 91, 114–16 (Tamás Zétényi ed., 1988) (indicating that people do not use the product rule naturally). In any event, in the legal system any human failing to conjoin properly would be offset by the human tendency to construct a story for the whole case instead of proceeding element-by-element. *Cf.* Tversky & Kahneman, *supra* note 90, at 19 (discussing biases that tend to ignore conjunction); Amos Tversky & Derek J. Koehler, *Support Theory: A Nonextensional Representation of Subjective Probability*, *in* HEURISTICS AND BIASES, *supra* note 90, at 441 (same).

supposedly deserving plaintiff (or to a defendant with multiple defenses almost proved):

> Consider a case involving three different legal theories and three different factual foundations. Plaintiffs deserve to win if one of the stories embodying one legal theory is true; defendants deserve to win only if all of their competing stories are true (for if this is false, one of the plaintiff's stories is true). For example, assume the plaintiff has alleged defective design, defective manufacture, and failure to warn theories. If the probability of each is .25, the "probability" of each not being true is .75, but, the probability of at least one being true is $1-.75^3=.58$, and perhaps plaintiff should win, even though the individual probabilities of each being false is .75.[124]

However, this paradox implodes under the force of fuzzy logic too. The MAX rule indicates that the plaintiff proved his case to only .25 and so should lose, just as the plaintiff does lose under current law.

Let me use this multiple-theory, or aggregation, paradox in trying further to explain what is not intuitive, and what is therefore difficult to explain. Imagine a claim A that is 25% likely to succeed and an independent claim B that is 40% likely to succeed. I am saying that

---

124   Allen & Jehl, supra note 110, at 939; *see* Massachusetts v. U.S. Dep't of Health & Human Servs., 682 F.3d 1 (1st Cir. 2012) (striking down DOMA on both equal protection and federalism grounds), *noted in* Mike Dorf, *Is the First Circuit's Opinion in the DOMA Case Insufficiently "Fuzzy"?*, DORF ON LAW (June 4, 2012, 12:30 AM), http://www.dorfonlaw.org.

A recent article attacks the law's general refusal to aggregate the probabilities of independent claims and defenses, even arguing for conviction upon the basis of a number of criminal offenses almost proved. Porat & Posner, *supra* note 70. *Compare* Alon Harel & Ariel Porat, *Aggregating Probabilities Across Cases: Criminal Responsibility for Unspecified Offenses*, 94 MINN. L. REV. 261, 261–62 (2009) ("Should a court convict a defendant for an unspecified offense if there is no reasonable doubt that he committed an offense, even though the prosecution cannot prove his guilt as to a particular offense beyond a reasonable doubt? Stated otherwise, is committing *an* offense sufficient for a conviction or must a prosecutor establish what this offense is to justify a conviction? This Article contends that, under certain conditions, a prosecutor should not have to establish the particular offense committed by a defendant—proof that the defendant committed *an* offense should be sufficient."), *with* Frederick Schauer & Richard Zeckhauser, *On the Degree of Confidence for Adverse Decisions*, 25 J. LEGAL STUD. 27, 41-47 (1996) (conceding the probability argument, but arguing that the criminal law still should not convict for reasons of abundant caution). My position is that the paradox that motivated their whole article does not exist. The law should not, and does not, aggregate. Their article's cited exceptions, where the law does seem to aggregate (for example, alternative or market share liability), actually constitute changes to the substantive law in a manner comparable to imposing strict liability, rather than resulting from an odd application of traditional probability to the standard of proof. *See* Kevin M. Clermont, *Aggregation of Probabilities and Illogic*, 47 GA. L. REV. 165 (2012).

the likelihood of (A OR B) is equal to that of the likeliest of A and B, that is, 40%. It is easy to say that I am being obtuse, because anyone can intuit that the probability of A and B's union must be higher. But the fact that the law says otherwise, and that a widely accepted logic system says otherwise, should give pause to intuition.

First, imagine a claim A for tort, where the defendant did acts that were 25% bad, and an independent claim B for contract, where the defendant did acts that constituted 40% of what would be an unquestionable breach. These are members of fuzzy sets. The breaching quality of the acts has no bearing on the tortiousness of the acts. Then we can say only that the defendant went 40% of the way toward liability. The 25% showing has no effect on the 40% showing. The "likelihood" of their union is 40%.

Second, imagine a ball A drawn from an urn with only 25% black balls among white balls, and a ball B drawn from another urn with 40% black balls. The odds that one of them is black (and let us say that black represents liability), when the drawn balls are both revealed, are 55% by De Morgan's rule. In a sense, the act of revealing the balls to be white or black affects the odds, because the balls must be either white or black and only one has to be black. If one turns up black, this takes the pressure off the other's being black.

Third, imagine a claim A for tort, where the defendant was actually 25% likely to have done the bad tortious acts alleged, and an independent claim B for contract, where the defendant was actually 40% likely to have done the bad breaching acts alleged. Assume that all available evidence would not change those numbers, which is what I mean by "actually." Is this situation more like the first or the second situation? All our intuitions, honed by life-long exposure to traditional probability theory, point us to the second. But the real world of uncertainty and imprecision makes the appropriate analogy the first situation. The likelihood that A exists has no effect on whether B exists. More complete evidence will not arrive to change the likelihoods—and we cannot pull off a veil to show what really happened, we will not get to see if the "ball" is truly black, we will never get to reduce the world to bivalence. It was the reduction to bivalence that affected the joint odds in the second situation. The likelihoods in the third situation will never be anything but 25% and 40%, unlike the drawn balls whose probabilities will change upon unveiling. We have nothing more than 25% of a claim and 40% of another claim, when only a provable claim justifies liability. If we can never convert the likelihood of a claim to one or zero, then all we can say is that the defendant is liable to a certain degree. Thus, when we can never know with certainty what happened, a likelihood of occurrence is not

different from a degree of misfeasance: now, likelihood of occurrence is not a traditional probability, it is a fuzzy set.

III.   ANALYZING BELIEFS

The ultimate focus on fuzzy provability pushed traditional probability farther into the background, so setting the stage for a shift of focus onto beliefs as being at the core of the standards of proof. This Part will introduce belief functions into the mix, in order better to represent how imperfect evidence keeps fact-finders from committing all of their belief. Then, this Part will use this theory to explain why the law's initial burden of production starts the fact-finders at point zero. While the key idea introduced heretofore has been multivalence, the key idea henceforth will be the non-additivity of beliefs.

A.   *Shafer's Belief Functions*

1.   Basics of Theory

I have already implicitly advocated that we treat the degree of S's truth, which is a degree of membership in the set of true facts, as a degree of belief in S as a true proposition. The broad version of the theory of belief functions will now give us a handle on how to manipulate such beliefs.[125] It will also provide us with a better mental image for representing indeterminacy.[126]

In fact-finding, I therefore contend, we should not ask how likely S is but rather how much we believe S to be a real-world truth based on the evidence, as well as how much we believe notS—while remaining conscious of indeterminacy and so recognizing that part of our belief will remain uncommitted. Beliefs can range anywhere between 0 and 1. If the belief in S is called Bel(S), then $0 \leq \text{Bel}(S) \leq 1$.

Consider belief function theory's treatment of a single factual hypothesis. Take as an example the issue of whether Katie is dead or alive, with S representing death. Although you have no tangible evidence, three witnesses said she is dead. One seems somewhat credible. But you think that another saw a different woman's body, which discounts the evidence of death but gives no support to her being alive. And you think that the third was lying as part of a cover-up of her escape from captivity, which is compatible with both S and notS

---

125   *See* SHAFER, *supra* note 23, at 35–37.

126   *See* Liping Liu & Ronald R. Yager, *Classic Works of the Dempster–Shafer Theory of Belief Functions: An Introduction, in* CLASSIC WORKS OF THE DEMPSTER–SHAFER THEORY OF BELIEF FUNCTIONS, *supra* note 55, at 1, 2–19 (recounting also the history of belief function theory).

and so gives some thin support to her being alive. In sum, this evidence supports your .5 belief that she is dead, or Bel(S). That evidence also supports your weaker belief that she is alive, with Bel(notS) coming in at .2. That is, Bel(notS) is not determined by the value of Bel(S). The remaining .3 is indeterminate, meaning she could be either alive or dead because the evidence is imperfect. The defects in evidence might be probative, affecting Bel(S) and Bel(notS); but the defects might be nonprobative, so that they just leave some belief uncommitted. (This example actually involves a so-called power set of four beliefs: S, notS, neither S nor notS, and either S or notS. The belief in the "null" of neither alive nor dead is set by definition to be 0. The belief in the "catchall" of either alive or dead is 1.0.)

← Plausibility of Katie's death →		
Bel(S)		Bel(notS)
0		1
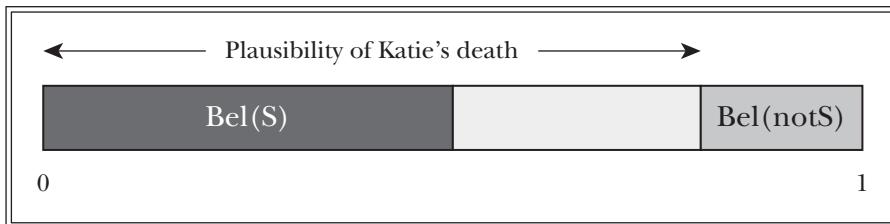
Belief is sometimes called the lower probability. Bel(S) is the extent to which you believe Katie to be dead. The upper probability bound represents "possibility" in Zadeh's terminology or "plausibility" in Shafer's.[127] It is the extent to which you think her being dead is plausible, that is, the sum of the affirmative belief plus the indeterminate belief. The plausibility that she is dead is .8, being .5 + .3. (A traditionally expressed probability of her being dead would fall somewhere within the range from the lower to the upper probability.) The plausibility that she is alive totals .5, being .2 + .3. Plausibility equals one minus the belief in the opposite.

Belief functions thus harness the idea of imprecise probability to capture indeterminacy. Although they can be used with ordinary expressions of probability, combining belief functions with fuzzy logic's degrees of truth and its operators makes an even bigger step toward understanding.[128] The resultant beliefs can be expressed, if

---

127   *See* Barnett, *supra* note 55, at 200–01 (providing a neat mental image for these bounds); A. P. Dempster, *Upper and Lower Probabilities Induced by a Multivalued Mapping*, 38 ANNALS MATHEMATICAL STAT. 325 (1967); L.A. Zadeh, *Fuzzy Sets as a Basis for a Theory of Possibility*, 1 FUZZY SETS & SYS. 3 (1978).

128   *See* Dubois & Prade, *supra* note 99, at 375 (arguing for the basic compatibility of the two approaches); John Yen, *Generalizing the Dempster–Shafer Theory to Fuzzy Sets, in* CLASSIC WORKS OF THE DEMPSTER–SHAFER THEORY OF BELIEF FUNCTIONS, *supra* note

expression is ever necessary, as coarsely gradated beliefs. In addition to the benefits of utilizing natural language, these terms capture the uncertainty and imprecision in determining the belief. Thus, in lieu of expressing beliefs in terms of decimals, one should use the coarse gradations of (1) slightest possibility, (2) reasonable possibility, (3) substantial possibility, (4) equipoise, (5) probability, (6) high probability, and (7) almost certainty.

In the end, the representation of findings in the form of beliefs captures the effect of imperfect evidence, which was a rallying cry of Baconian theorists.[129] The shift from probability to belief is also a slight nod to the civil-law emphasis on inner belief as captured by its *intime conviction* standard,[130] and to the frequent *cris de coeur* of theorists who lament any intrusion of probabilistic mathematics into the very human process of proof.[131] Finally, belief functions can make a contribution to understanding law independently of fuzzy theory, as I shall try to show.

55, at 529 (showing how to form beliefs about membership in a fuzzy set); *cf.* SCHUM, *supra* note 22, at 266–69 (observing that one can fuzzify belief functions).

  129 *See* COHEN, *supra* note 102, at 49–57, 245–64 (developing, as an alternative to Pascalian (or mathematicist) probability, a Baconian (or inductive) theory of probability). Baconian theory tries to look not only at the evidence presented, but also at the evidence not available. It makes *evidential completeness* a key criterion, and thereby stresses an important concern.

  130 *See* Clermont & Sherwin, *supra* note 32; Kevin M. Clermont, *Standards of Proof in Japan and the United States*, 37 CORNELL INT'L L.J. 263 (2004); Clermont, *supra* note 31; Wright, *supra* note 100; Richard W. Wright, *Proving Causation: Probability Versus Belief, in* PERSPECTIVES ON CAUSATION 195 (Richard Goldberg ed., 2011). With their emphasis on "conviction" in the *intime conviction* standard, the civil-law countries signal their devotion to belief, albeit a belief seemingly built upon a binary world view (and perhaps a belief compared to an absolute threshold inherited from the criminal model). Such an approach fit better with an inquisitorial model than it did with an adversarial model, allowing it to persist for centuries. But its survival until today may rest instead on the civil-law system's desire to enhance the appearance of legitimacy.

  131 *See, e.g.*, Jaffee, *supra* note 45, at 934–51 (attacking the use of probability in analyzing proof); Tribe, *supra* note 2 (writing the classic version of the lament, in which Professor Tribe stressed not only the risk of misuse of mathematical techniques, including inaccurate meshing of numerical proof with soft or unquantifiable variables, but also the undercutting of society's values, including the dehumanization of the legal process); Adrian A.S. Zuckerman, *Law, Fact or Justice?*, 66 B.U. L. REV. 487, 508 (1986) (arguing that probabilistic assessment diminishes "the hope of seeing justice supervene in individual trials," while seeing fact-finding as an individualized but value-laden process).

2. Negation Operator

By traditional probability theory, the probability of a hypothesis's negation equals 1 minus the probability of the hypothesis. If Katie is 60% likely dead, she is 40% likely alive.

Under the scheme of belief functions, Bel(S) and Bel(notS) do not necessarily add to 1, because normally some belief remains uncommitted. Thus, for Katie, Bel(S)=.5 and Bel(notS)=.2, so the sum of determinate beliefs adds to .7. We are now squarely in the realm of non-additive beliefs.

The complement of Bel(S) equals (1 - Bel(S)), but it gives the plausibility of notS, not the belief in notS. Indeed, the plausibility of notS equals (Bel(notS) + uncommitted belief). Hence, there is a big difference between the complement and the belief in the negation: the difference is the uncommitted belief. Belief function theory thus utilizes the very useful distinction between a lack of belief and a disbelief. After all, disbelief and lack of belief are entirely different states of mind.[132]

3. Lack of Proof

Traditional probability encounters legendary difficulties with a state of ignorance.[133] The reason is that it cannot distinguish between lack of belief and disbelief. In classical terms, S=0 means that S is impossible. And it means that notS is certain. No amount of evidence could alter an impossibility or a certainty into a possibility under Bayes' theorem.[134] As a way out, probabilists sometimes assert that the ignorant inquirer should start in the middle where the probabilities of S and notS are both 50%. But this trick does not accord with the actual probabilities, and it produces inconsistencies when there are more than two hypotheses in play.[135]

Meanwhile, one of the great strengths of belief function theory is that it well represents a state of ignorance.[136] An inquirer, if ignorant, starts at zero, not at a 50% belief. When Bel(S)=0, it does not mean

---

132   *See* Bellman & Giertz, *supra* note 77, at 155–56 (showing that negation can have multiple meanings).

133   *See, e.g.*, Richard Lempert, *The New Evidence Scholarship: Analyzing the Process of Proof*, 66 B.U. L. REV. 439, 462–67 & n.60 (1986) (noting that employing 1:1 as the appropriate odds for someone who is ignorant of the true facts can cause many problems).

134   *See* Brilmayer, *supra* note 5, at 686–88.

135   *See supra* note 45 and accompanying text.

136   *See* SHAFER, *supra* note 23, at 22–24 (exploring the role of the "representation of ignorance" in belief functions).

that S is so highly unlikely as to be impossible. It means there is no evidence in support. Accordingly, the inquirer starts out with everything indeterminate, because the lack of evidence makes one withhold all of one's belief. Although Bel(S)=0, Bel(notS) equals zero too. The uncommitted belief is the entirety or 1, meaning that S is completely plausible, as is notS. In other words, the inquirer does not believe or disbelieve S. Belief function theory thus utilizes the very useful distinction between disproof and lack of proof.

B. Legal Application: Burden of Production

Let me start with some background on how the law has traditionally viewed the burden of proof, say, in a jury trial. The burden of proof dictates who must produce evidence and ultimately persuade the fact-finder on which elements of the case. Burden of proof thus encompasses two concepts: burden of production and burden of persuasion. The burden of production might require either party at a given time during trial to produce evidence on an element or suffer the judge's adverse determination on that element; one party has the initial burden of production on any particular element, but that burden may shift during the trial if that party produces certain kinds or strengths of evidence. The burden of persuasion requires a certain party ultimately to persuade the fact-finder of the truth of an element or suffer adverse determination on that element.

Imagine a single disputed issue of typical fact on which the plaintiff bears the initial burden of production and the burden of persuasion. Then imagine a grid representing the judge's disagreement with a potential verdict for the plaintiff, or equivalently the judge's view of likelihood of error in such a verdict, with disagreement or likelihood decreasing from one on the left to zero on the right.[137] It is important to realize that this diagram represents the likelihood of jury error in finding that the disputed fact exists, not the judge's view of the evidential likelihood that the disputed fact exists. In other words, this diagram represents the judge's thought process in externally overseeing the jury that acts as fact-finder, not the judge's thought process as if the judge were finding facts. Alternatively stated, this diagram represents the burden of production, not the burden of persuasion.

---

137   *See* 9 John H. Wigmore, Evidence § 2487 (James H. Chadbourn rev., 1981); *cf.* John T. McNaughton, *Burden of Production of Evidence: A Function of a Burden of Persuasion*, 68 Harv. L. Rev. 1382 (1955) (offering alternative diagrams).

Judge	Jury	Judge
X		Y
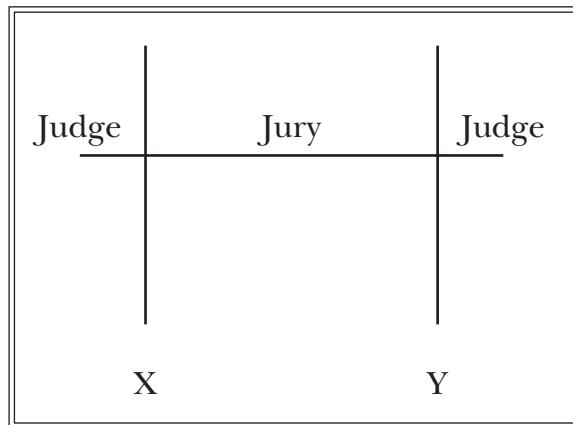
   The plaintiff in the imagined case starts at the left of the diagram. If he presents no evidence, the judge would ordinarily grant a motion for judgment as a matter of law against him. He is consequently bound to go forward with his evidence until he satisfies the judge that a reasonable jury would be warranted in finding for him. That is, he must get to line X in order to make a jury question of the imagined single issue of fact, doing so by presenting evidence. The plaintiff's getting to or beyond line X means that although the judge might still disagree with a verdict for the plaintiff, the judge thinks a reasonable jury could find that the plaintiff sustained his persuasion-burden, and therefore the judge will hold that the plaintiff sustained his production-burden. If the plaintiff does not get to line X, that means that the judge would so vehemently disagree with a verdict for the plaintiff as to consider the jury irrational, and so the judge can grant the motion for judgment as a matter of law. Line X, again, represents the judge's view on the limit of rationality in the jury's finding for the plaintiff, rather than the judge's view of the evidential likelihood that the disputed fact exists. For example, if the judge disbelieved all of the plaintiff's abundant evidence, but still acknowledged that a reasonable jury could believe it, then the judge should rule that the plaintiff has carried his production-burden, because a reasonable jury could conclude that the plaintiff sustained his persuasion-burden.
   This diagrammatic scheme works pretty well to represent the law's approach. Moreover, the diagram helps in understanding other concepts and special rules. A *permissive inference* (and res ipsa loquitur is one in the view of most courts[138]) describes an inference that a jury

---

138   *See* John Farley Thorne III, Comment, *Mathematics, Fuzzy Negligence, and the Logic of Res Ipsa Loquitur*, 75 Nw. U. L. Rev. 147 (1980) (justifying the res ipsa loquitur doctrine by use of fuzzy logic).

is authorized but not required to draw from certain evidence; in other words, the inference satisfies the plaintiff's production-burden by getting the case to line X, although not beyond line Y. A true *presumption* (such as the presumption against suicide as the cause of death) shifts the burden of production to the opponent after the introduction of the evidential premise; in other words, the presumption puts the case to the right of line Y and so requires the jury to find the presumed fact, unless the opponent introduces enough evidence to carry her production-burden and push the case at least back into the jury zone between Y and X.[139]

Among special rules, certain kinds of evidence will not satisfy an initial burden of production. To satisfy that burden, the burdened party cannot rely on the opponent's failure to testify,[140] on mere disbelief of the opposing testimony,[141] or on demeanor evidence drawn from the opponent's testimony.[142] Similarly, naked statistical evidence normally will not satisfy the initial burden of production.[143] However, any of these kinds of evidence is perfectly proper to introduce as a supplement to positive evidence that satisfies the initial burden of production.[144] The idea behind these special rules is that they are necessary to protect the notion of an initial burden of production, which serves to facilitate early termination of weak claims or defenses, to safeguard against irrational error, and to effectuate other process and outcome values.[145] In the absence of these special rules, any bur-

---

139   *See* Fed. R. Evid. 301.

140   *See* Stimpson v. Hunter, 125 N.E. 155, 157 (Mass. 1919) ("[T]he failure of the defendant and his son to testify although present in court was not equivalent to affirmative proof of facts necessary to maintain the action.").

141   *See* Cruzan v. N.Y. Cent. & Hudson River R.R. Co., 116 N.E. 879, 880 (Mass. 1917) ("Mere disbelief of denials of facts which must be proved is not the equivalent of affirmative evidence in support of those facts.").

142   *See* Dyer v. MacDougall, 201 F.2d 265, 269 (2d Cir. 1952) (holding that although demeanor evidence is probative, it does not suffice to escape a directed verdict).

143   *See* Guenther v. Armstrong Rubber Co., 406 F.2d 1315, 1318 (3d Cir. 1969) (dictum) (saying, in a case where the plaintiff had been injured by an exploding tire, that a 75 to 80% chance it came from the defendant manufacturer was not enough for the case to go to the jury). For a more complete consideration of statistical evidence and its ultimately non-paradoxical nature, see Richard H. Field, Benjamin Kaplan & Kevin M. Clermont, Materials for a Basic Course in Civil Procedure 1352–56 (10th ed. 2010) (explaining how a fact-finder converts statistical evidence into a belief).

144   *See* Baxter v. Palmigiano, 425 U.S. 308, 316–20 (1976) (treating failure to testify as sufficient evidence).

145   *See* Robert S. Summers, *Evaluating and Improving Legal Processes—A Plea for "Process Values,"* 60 Cornell L. Rev. 1, 4 (1974) (arguing that a legal process can be both a

dened party could produce enough evidence to reach the jury, this evidence possibly being merely in the form of silence, disbelief, demeanor, or general statistics (such as that the defendant manufactured 60% of the supply of the injury-causing device of unknown provenance). Perhaps we harbor a special fear of the jury's mishandling of such evidence when undiluted by other admitted evidence and consequently rendering an unreasoned verdict for the proponent based either on prejudice without regard to the evidence or on undue deference to such bewildering evidence. To avoid such an outcome, and to ensure that the burden of production means something, the judge should require sufficient evidence of other kinds. Once the proponent clears that hurdle, the tribunal should allow the feared evidence its probative effect.

At first glance, this whole accepted scheme seems fairly compatible with traditional probability. One diagrammatic qualification coming from the new logic would be that representing the judge's view of jury error as a fuzzy interval rather than a point would better capture reality.

But the biggest difficulty for traditional probability is fixing the starting point. The probabilist might assume that when you know nothing, the rational starting point is 50% (thus, many a Bayesian would make 50% the initial prior probability). Indeed, some experimental evidence indicates that lay people do tend to start at 50%.[146] Then, if the plaintiff offers a feather's weight of evidence, he would thereby carry not only his burden of production but also his burden of persuasion.

The real-life judge, however, hands only defeat to the plaintiff with nothing more than a feather's weight of evidence, and does so by summary means. Why is that? The law says that we should start not at 50% but at the far left, and to get to X requires more than a feather's weight. The proper representation of lack of proof is zero belief in the plaintiff's position, but also zero belief in the defendant's position. The full range of belief is properly uncommitted. That insight makes sense of the notion of the burden of production. It also suggests that, in starting at zero belief, the law is proceeding by belief function theory.

---

means to good results and a means to serve process values such as "participatory governance, procedural rationality, and humaneness").

146   *See* Anne W. Martin & David A. Schum, *Quantifying Burdens of Proof: A Likelihood Ratio Approach*, 27 JURIMETRICS J. 383, 390–93 (1987) (surveying a small sample of students for their odds of guilt used as the prior probability, which turned out to be 1:1 or 50%).

IV. APPLYING STANDARDS

This Part will introduce the idea of comparing belief and disbelief of a fact, which the fact-finder would do after putting any indeterminate belief aside. Then, this Part will demonstrate how the law already conceives of its three standards of proof as different ways of so comparing belief and disbelief.

A. Comparison of Beliefs

My conceptualization has thus far led me to think that the law should not and does not employ the traditional academic view of the proof process resting on a two-valued logical approach. Fact-finders instead determine their beliefs as fuzzy degrees of real-world truth based on the evidence, just as the law expects of them. Eventually they end up with Bel(S) and Bel(notS), falling between 0 and 1, but not necessarily adding to 1. What then do they do?

So, finally, I come to the matter of applying a standard of decision. The law dictates that fact-finders decide by subjecting their fuzzy beliefs to a standard of proof in order to come to an unambiguous output. That is, at this point the law forces fact-finders back into what looks like a two-valued logic, by forcing them to decide for one party or the other.[147] Such disambiguation is not a practice unique to law. All fuzzy computer programs end with a step that produces an unambiguous output, a step called *defuzzification*.[148]

Application of a standard of proof is a different step from evidential argument; the academic disputes as to standards do not overlap with the disputes over how to assess, conjoin, and analyze evidence.[149]

---

147 A separable question is whether the law should instead deliver partial relief following partial proof. *See* ENDICOTT, *supra* note 10, at 72–74. The answer is not obviously affirmative. *See* David Kaye, *The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation*, 1982 AM. B. FOUND. RES. J. 487 (showing current law's economic superiority to an expected-value approach that would award damages proportional to probabilistic certainty). After cataloging the prevalence of fuzzy concepts in law, Professor Katz concludes that law is correct to draw lines in the fuzz and so separate all-or-nothing remedies; he argues that law must establish discontinuities in order to retain its rational coherence. *See* KATZ, *supra* note 67, at 157–81. Most significantly, I add that partial relief would have to contend with the difficulty that the system does not, and logically could not, charge the fact-finder to find upon imperfect evidence the degree to which the plaintiff is right, as I next develop.

148 *See* KOSKO, *supra* note 10, at 172 (describing the step in fuzzy computer systems).

149 *See* Peter Tillers & Jonathan Gottfried, *Case Comment—United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond a Reasonable Doubt Is Unquantifiable?*, 5 LAW PROBABILITY & RISK 135, 142

Psychologists have contributed almost nothing here,[150] leaving the dispute to logicians so far.

As to the psychology involved, I assume only that people, if told to do so, can apply a simple standard of proof imposed by law. Similarly, in the absence of studies to the contrary, I believe jurors and others will try to do so. Thus, it matters what the law says about standards.

On the logic front, I contend that speaking in terms of two-valued logic tends to mislead on standards, just as it does elsewhere. Admittedly, the determined theorist could pursue the two-valued image of traditional probability. Then the ultimate task of applying a standard of proof would unavoidably involve placement on a scale of likelihood.[151]

A better understanding of standards of proof would result from thinking in terms of many-valued logic and belief functions, however. Even though decision-making requires converting from a many-valued logic to an output that sounds two-valued, the law does not need to require enough evidence to make the fact more likely than 50% or whatever. The path to decision might involve only comparing Bel(S) and Bel(notS) while ignoring the indeterminate belief. All the fact-finder need do is compare the strengths of belief and disbelief. By requiring only a comparison, belief functions would never require placement on a scale of likelihood.[152]

B.   Legal Application: Burden of Persuasion

1.   Traditional View

In going from discussing the burden of production to explaining the academic view of the burden of persuasion, I need to use a different diagram, one that represents the internal thought process of the


(2006) (observing the difference between processing evidence and applying standards of proof); *cf.* Pardo, *supra* note 118, at 8 (calling these two stages the micro-level and the macro-level of proof). Two Belgian scholars helpfully elaborated belief functions to distinguish the formulation of beliefs during a "credal" stage (from the Latin for believe) and decision-making during the "pignistic" stage (from the Latin for a bet). Philippe Smets & Robert Kennes, *The Transferable Belief Model, in* Classic Works of the Dempster–Shafer Theory of Belief Functions, *supra* note 55, at 693.

150   *See* Clermont, *supra* note 31, at 477, 485 (recounting the state of psychological knowledge on standards of proof).

151   *See, e.g.,* Brown v. Bowen, 847 F.2d 342, 345 (7th Cir. 1988) ("[T]he trier of fact rules for the plaintiff if it thinks the chance greater than 0.5 that the plaintiff is in the right.").

152   *See* Smith, *supra* note 71, at 503 (describing a betting scheme based on such comparison). For the mathematics involved, see Smets & Kennes, *supra* note 149, at 703–11.

fact-finder in ultimately weighing the evidence. The grid now measures the fact-finder's view of the evidential likelihood that the disputed fact exists, with likelihood increasing from 0% on the left to 100% on the right.
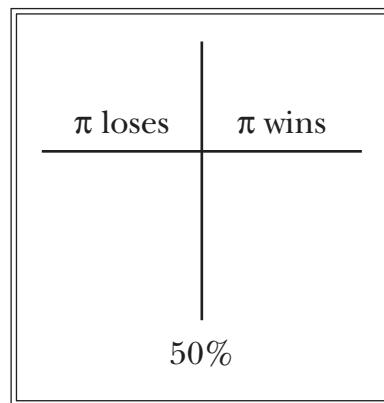
The plaintiff in an imagined civil trial again starts at the left. By presenting evidence on the issue, he must get beyond the midpoint to win. That is, he must show that it is more likely than not that the disputed fact exists. If, after the plaintiff has given his best shot, the fact-finder thinks that he has not passed the 50% line, then the fact-finder should decide for the defendant.

A necessary qualification is that even under this traditional view, this diagram serves mainly as an impetus to thinking about these matters, rather than as a source of definitive statements thereon. For example, the diagram does not mean that a 50% line exists in reality. The psychological truth is that equipoise is more of a zone, or range of probabilities, than a line. A range of evidential states may strike the fact-finder as evenly balanced.[153]

That equipoise is a zone means that the burden of persuasion will affect many more cases than those in which the conflicting evidence results precisely in a dead heat. The fact-finder will rely on the burden of persuasion more often than one might imagine. Also, how the law frames an issue—whether the plaintiff or the defendant bears the

---

153   *See* United States *ex rel.* Bilyew v. Franzen, 686 F.2d 1238, 1248 (7th Cir. 1982) (stressing importance of the persuasion burden and observing that "a judge or a jury can experience only a small, finite number of degrees of certainty . . . . Thus cases when the evidence . . . seem[s] in balance are not unique among some infinite variety of evidentiary balances, but instead are among a much smaller number of [ranges of] possibilities that may be perceived by the fact-finder."); Clermont, *supra* note 20, at 1119 n.13, 1122 n.36, 1147–48; Cohen, *Conceptualizing Proof, supra* note 30, at 90–91; Cohen, *Confidence in Probability, supra* note 30, at 418–19.

brunt of nonpersuasion of a fact, that is, whether the plaintiff or the defendant appears to start from zero—matters.[154] An anchoring heuristic lowers the willingness of the fact-finder to determine that the burdened party has prevailed, because people fail to adjust fully from a given starting point, even if arbitrarily set.[155] In sum, the burden of persuasion is not a mere tiebreaker, which explains why lawyers and judges fight and suffer over it in practice.

Again, this diagrammatic representation of the traditionally viewed burden of persuasion appears fairly compatible with traditional probability. But having to draw a fat 50% line encourages a reconsideration of the proof standards. And that reconsideration leads to reformulating those standards to reflect the role of the new logic. The conclusion will be that this diagram for the burden of persuasion is fundamentally misleading. The diagramed view thus needs redrawing rather than mere refinement. The law does not and should not conform to the traditional academic view.

2.   Reformulated View

    a.   Current Standards

The law has settled on three standards of proof that apply in different circumstances: (1) The standard of *preponderance of the evidence* translates into more-likely-than-not. It is the usual standard in civil litigation, but it appears throughout law. (2) Next comes the intermediate standard or standards, often grouped under the banner of *clear and convincing evidence* and roughly translated as much-more-likely-than-not. These variously phrased but equivalently applied standards govern on certain issues in special situations, such as when terminating parental rights.[156] (3) The standard of *proof beyond a reasonable doubt* means proof to a virtual-certainty. It very rarely prevails outside criminal law.[157]

---

154   *See* Eyal Zamir & Ilana Ritov, *Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation*, 41 J. LEGAL STUD. 165, 197 n.23 (2012).

155   *See* Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE (n.s.) 1124, 1128 (1974). An example of the anchoring heuristic comes from a study involving subjects asked to estimate quickly, without paper and pencil, the product of 8 x 7 x 6 x 5 x 4 x 3 x 2 x 1, while another group faced 1 x 2 x 3 x 4 x 5 x 6 x 7 x 8; the first group's median estimate was 2250, while the other's was 512; the correct answer is 40,320. *Id.*

156   *See* Santosky v. Kramer, 455 U.S. 745, 757 (1982) ("[T]his Court never has approved a case-by-case determination of the proper *standard of proof* for a given proceeding.").

157   *See generally* Clermont, *supra* note 20, at 1118–21 (listing the standards of proof and their relative applications to different areas of the law).

b.   Relative-Plausibility Theory

The insightful relative-plausibility theory of Professor Ron Allen shows a nontraditional embrace of relative judgment, in preference to our weaker skills at absolute judgment of likelihood.[158] He builds on the story model of evidence-processing to produce another theoretical brand of holism. The relative-plausibility theory posits that the fact-finder constructs the story (or stories) that the plaintiff is spinning and another story (or stories) that the defendant is spinning. The fact-finder then compares the two stories (or collections of stories) and gives victory to the plaintiff if the plaintiff's version is more plausible than the defendant's.[159] This choice between alternative competing narratives is largely an ordinal process rather than a cardinal one.

Allen's ordinal comparison cannot easily explain standards of proof higher or lower than preponderance of the evidence.[160] Its

---

158   *See* Allen & Jehl, *supra* note 110, at 929–43 (summarizing Allen's previous work on the theory); *see also* Edward K. Cheng, *Reconceptualizing the Burden of Proof*, 122 YALE L.J. (forthcoming 2013) (manuscript at 3–4, 6), *available at* http://ssrn.com/abstract =2087254 (arguing that statisticians perform hypothesis testing by comparison, so that "evidence scholars need only let go of their love for $p > 0.5$"; but incorrectly assuming that theory calls for comparing the plaintiff's story to each of the defendant's stories "separately, not simultaneously"); *supra* note 81 and *infra* note 166.

159   The *weight of the evidence* methodology in science is a similar approach, as is the *differential diagnosis* approach in medicine that diagnoses by successively eliminating likely causes of a medical condition to reveal the best explanation. *See* Milward v. Acuity Specialty Prods. Grp., Inc., 639 F.3d 11, 18 (1st Cir. 2011) ("The scientist [when admitting expert evidence based on the weight of the evidence approach] must (1) identify an association between an exposure and a disease, (2) consider a range of plausible explanations for the association, (3) rank the rival explanations according to their plausibility, (4) seek additional evidence to separate the more plausible from the less plausible explanations, (5) consider all of the relevant available evidence, and (6) integrate the evidence using professional judgment to come to a conclusion about the best explanation."); Westberry v. Gislaved Gummi AB, 178 F.3d 257 (4th Cir. 1999) (admitting expert evidence based on differential diagnosis). These methods involve consideration and analysis of alternative explanations to get the one that best explains the evidence, a mode of reasoning called *inference to the best explanation*. Allen is drifting in his thinking in this direction. *See* Ronald J. Allen & Michael S. Pardo, *Juridical Proof and the Best Explanation*, 27 LAW & PHIL. 223, 226 (2008) (providing an account of the "abductive reasoning process of inference to the best explanation"). However, Larry Laudan, *Strange Bedfellows: Inference to the Best Explanation and the Criminal Standard of Proof*, 11 INT'L J. EVIDENCE & PROOF 292 (2007), powerfully demonstrates that inference to the best explanation holds little additional promise of explaining or illuminating standards of proof.

160   *See* Richard D. Friedman, *"E" Is for Eclectic: Multiple Perspectives on Evidence*, 87 VA. L. REV. 2029, 2046–47 (2001); *cf.* Clermont, *supra* note 20, at 1119–20, 1122–26 (discussing not only the standard of clear and convincing evidence, but also procedure's lower standards including slightest, reasonable, and substantial possibilities).

more obvious, and admitted,[161] difficulty is that it does not track well what the law tells its fact-finders about how to proceed, and it diverges from the law by compelling the non-burdened party to choose and formulate a competing version of the truth. Finally, it comes with baggage, such as requiring acceptance of the story model.[162]

c. Reformulated Standards

Consider what else preponderance of the evidence, or its translation of more-likely-than-not, could mean in a comparative sense.

One could compare the proof to some threshold. Although one could say that the proof must exceed 50%, this formulation does not accord with the import of real cases. The law does not require the completeness of proof that would be necessary to get a belief above 50%. The law is willing to rest decisions on the evidence presented.[163]

The law does not inquire which side has the stronger evidence, however. It looks instead to belief in the burdened party's position.[164] Although one could measure the belief against some absolute measure, say, requiring that Bel(S) exceed 50%, the better approach is to invoke the more powerful human ability of relative judgment by comparing beliefs. One could compare Bel(S) relative to Bel(notS).[165]

But cf. Ronald J. Allen, *The Nature of Juridical Proof*, 13 Cardozo L. Rev. 373, 413 (1991) (attempting to explain the beyond-a-reasonable-doubt standard as not being satisfied if the fact-finder "concludes that there is a plausible scenario consistent with innocence," while admitting that the clear-and-convincing standard is "troublesome" under his theory because it seems cardinal); Ronald J. Allen & Brian Leiter, *Naturalized Epistemology and the Law of Evidence*, 87 Va. L. Rev. 1491, 1528 (2001) ("[T]he prosecution must provide a plausible account of guilt and show that there is no plausible account of innocence.").

161   *See* Ronald J. Allen, Standards of Proof and the Limits of Legal Analysis 14 (May 25–26, 2011) (unpublished conference paper), *available at* http://ssrn.com/abstract=1830344.

162   *See* Craig R. Callen, Commentary, *Kicking Rocks with Dr. Johnson: A Comment on Professor Allen's Theory*, 13 Cardozo L. Rev. 423 (1991) (arguing that for the purposes of the study of evidence and fact-finding, the insights from cognitive science may be more far-reaching than Professor Allen suggested in *The Nature of Juridical Proof*).

163   *See* Laudan, *supra* note 159, at 304–05 ("The trier of fact cannot say, 'Although plaintiff's case is stronger than defendant's, I will reach no verdict since neither party has a frightfully good story to tell.' Under current rules, if the plaintiff has a better story than the defendant, he must win the suit, even when his theory of the case fails to satisfy the strictures required to qualify his theory as the best explanation.").

164   *See* McBaine, *supra* note 79, at 248–49.

165   *See* Cohen, *supra* note 102, at 255 ("The cardinal question to be settled by the trier of fact may always be construed as this: on the facts before the court, is the conclusion to be proved by the plaintiff more inductively probable than its negation?").

In comparing them, Bel(notS) is the belief in the negation of S, not the complement of Bel(S). It represents how much the fact-finder actively disbelieves S, the fact in dispute. The comparison thus should look at actual belief in S and actual disbelief of S.

If you were to work with only those two beliefs, and discard the indeterminate belief, the most obvious course in civil cases would be to say that the burdened party should win if and only if Bel(S) > Bel(notS). You would decide for the plaintiff if Bel(S) exceeds Bel(notS), but decide for the defendant if Bel(S) does not exceed Bel(notS).

This comparative approach to the civil standard of proof does not mean that the non-burdened party needs to formulate a competing version of the truth, other than negation. A belief in the falsity of the burdened party's version of the truth may develop naturally in the course of trial. It could arise even upon hearing only the burdened party's evidence. The non-burdened party's evidence, if any, should contribute to raising Bel(notS).

Relatedly, the non-burdened party need not fight imaginary fights. Some scholars worry that looking at negation puts the burdened party in the impossible situation of disproving every alternative possibility.[166] But that worry comes from confusing lack of belief with disbelief. Disbelieving S entails the degree to which the fact-finder

---

166   *See, e.g.*, Michael S. Pardo, *Second–Order Proof Rules*, 61 FLA. L. REV. 1083 (2009) (speaking of the comparison imposed by more likely than not, but using "negation" in the sense of the complement of Bel(S)). Pardo explains that the comparison

> might mean the likelihood of the plaintiff's factual allegations versus the negation of those allegations, or it might mean the likelihood of the plaintiff's allegations versus the likelihood of the defendant's alternative allegations. The first interpretation appears to better fit the instructions, but it fails . . . . If the plaintiff must prove that some fact, *X*, is more probable than its negation, *not-X*, then the plaintiff should have to show not only the probability that the state of the world is such that *X* is true, but also the probability of every other possible state of the world in which *X* is not true. This would mean that in order to prevail, plaintiffs would have to disprove (or demonstrate the low likelihood of) each of the virtually limitless number of ways the world could have been at the relevant time. This would be a virtually impossible task, and thus, absent conclusive proof, plaintiffs would lose. This would plainly be inconsistent with the goals of the preponderance rule, and thus some comparison with the defendant's case is necessary.
>
> In order to facilitate the goals of the preponderance rule, the plaintiff ought to prevail whenever the likelihood of his allegations exceeds that of the defendant's.

Id. at 1093–94 (footnotes omitted). The difficulty for theorists who compare the plaintiff's story to the defendant's story, rather than to all versions of non-liability, is that plaintiffs will recover more often than normatively desirable. Realization of this

thinks S is false. The mere possibility of other states of the world in which S is not true go into the uncommitted belief, not into Bel(notS); recall that the "plausibility" of notS equals Bel(notS) plus the uncommitted belief; again, the degree of believing that Katie is not dead, or actually alive, is quite different from envisaging the chance that she is possibly alive. The proposed comparison involves the belief in notS, and does not involve the plausibility of notS.

Now, as to the other two standards of proof, clear and convincing evidence should mean Bel(S) >> Bel(notS).[167] This standard would not be that difficult to apply. We are quite used to such a standard of being clearly convinced, in life and in law. Judges apply it on a motion for a new trial based on the verdict's being against the weight of the evidence.[168] Appellate courts use it in reviewing judge-found facts.[169] Those standards of decision mean that it is not enough to disagree with the jury or the judge; the reviewer must think there was a serious error.

However, the cases do not make very evident what clear and convincing means. Alternatively, or perhaps additionally, it imposes a requirement about the completeness of evidence. It may require admission of enough evidence to reduce uncommitted belief to the point that Bel(S) exceeds the plausibility of notS. I am open to those viewpoints, but unconvinced so far.

As to proof beyond a reasonable doubt, it is different in kind. It must mean more than Bel(S) >> Bel(notS). Placing separate demands on Bel(notS) and Bel(S), it should mean that no reasonable doubt persists and that no great uncommitted belief remains.[170]

---

difficulty leads some of the theorists to argue that the aim of the system is not truth but, say, acceptability of decision. *See* Nesson, *supra* note 113.

167   *See* McBaine, *supra* note 79, at 263 (proposing an instruction to the effect that "the probability that they are true or exist is substantially greater than the probability that they are false or do not exist"); Edmund M. Morgan, *Instructing the Jury upon Presumptions and Burden of Proof*, 47 HARV. L. REV. 59, 67 (1933) ("[I]f the judge charges that the burden is upon a party to prove a proposition by clear and convincing evidence, . . . it requires the jury to be convinced . . . that its truth is much more probable than its falsity . . . ."); c*f.* Laudan, *supra* note 159, at 299–300 (discussing attempts to append such a notion to the approach of inference to the best explanation).

168   *See* Clermont, *supra* note 20, at 1126–28, 1152–56 (describing new-trial practice).

169   *See id.* at 1128–30 (describing appellate review).

170   *See* Allen & Leiter, *supra* note 160, at 1528 ("[T]he prosecution must provide a plausible account of guilt and show that there is no plausible account of innocence."); McBaine, *supra* note 79, at 266 ("A reasonable doubt is a doubt which exists[ ] when . . . you cannot honestly say that it is almost certain that the defendant did the acts which he is charged to have done."); *cf.* Laudan, *supra* note 159, at 300–02 (discussing

No reasonable doubt means that no reasonable person could hold Bel(notS) > 0. On the view that anything is possible, zero as a coarsely gradated degree of belief equates to a "slightest possibility."[171] Therefore, Bel(notS) > 0 refers to a step up from the slightest possibility of innocence. No reasonable fact-finder should see a "reasonable possibility" of innocence. In other words, for a conviction the prosecutor must show that no reasonable possibility of innocence exists.

No great uncommitted belief reflects the idea that Bel(S) cannot be weak, measured in an absolute sense. We do not want to convict when, although there is some evidence of guilt, we really do not know what happened. The belief in guilt must outweigh all alternative possibilities, including fanciful ones. The *belief* in guilt must exceed the *plausibility* of innocence, so that Bel(S) > .50. Given the usual limits on available evidence, achieving such a high degree of absolute belief represents a demanding standard.[172]

d.   Compatibility of Reformulated and Current Standards

A reader always entertains the temptation, upon seeing what looks like a plea for reconceptualization, to dismiss it as a pie-in-the-

---

attempts to append such notions to the approach of inference to the best explanation).

171   *See* Michael J. Saks & Robert F. Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 LAW & SOC'Y REV. 123, 126 (1981) ("Most legal decision making, like that in many other areas of complex activity, is done under conditions of uncertainty."). One can play with the notion of uncertainty, of course. For example, "to be certain of uncertainty . . . is to be certain of at least one thing." Milton Dawes, *Multiordinality: A Point of View*, ETC, Summer 1986, at 128, 131; *cf.* Neal Gabler, *The Elusive Big Idea*, N.Y. TIMES, Aug. 14, 2011 (Sunday Review), at 1, *available at* http://www.nytimes.com/2011/08/14/opinion/sunday/the-elusive-big-idea.html (announcing as the big idea that we are in a post-idea era). Philosophers can do more than play with the notion. *See, e.g.*, Daniel Greco, *Probability and Prodigality*, 4 OXFORD STUD. EPISTEMOLOGY (forthcoming 2013), *available at* http://web.mit.edu/dlgreco/www/ProbAndProd.pdf (objecting to the view that what we know has a probability of one); Gary Lawson, *Proving the Law*, 86 NW. U. L. REV. 859, 871–74 (1992) (discussing possible Cartesian arguments regarding uncertainty). For some propositions in closed systems, uncertainty will seem rather thin, except in the metaphysical sense that nothing is absolutely certain. But in almost all circumstances calling for the application of law, out in the real world, uncertainty will be a palpable concern. *See* TWINING, *supra* note 1, at 104 (rebutting "the myth of certainty").

172   On justifying what still may seem to be a low threshold, Bel(S) > .50, see Ronald J. Allen & Larry Laudan, *Deadly Dilemmas*, 41 TEX. TECH. L. REV. 65 (2008); Larry Laudan & Harry D. Saunders, *Re-thinking the Criminal Standard of Proof: Seeking Consensus About the Utilities of Trial Outcomes*, 7 INT'L COMMENT. ON EVIDENCE iss. 2, art. 1 (2009).

sky academic musing. When the reconceptualization involves the standards of proof, the specialists have the added temptation of dismissing it as another of the common anti-probabilist rants or pro-probabilist paeans. After all, if my view were a sound one, someone would have come up with it before. So I hasten to undercut my contribution by stressing that my ideas are not that new. I am trying little more than to explain what the law has been doing all along.

The easiest way to grasp the lack of newness is to picture an alternative fashion of converting from fuzzy beliefs back into a two-valued output. Picture a normalization process of disregarding the indeterminate beliefs and scaling Bel(S) and Bel(notS) up proportionately so that they add to one. Call the recalculations b(S) and b(notS). If Bel(S)=.50 and Bel(notS)=.20, then b(S)=.71 and b(notS)=.29. These new numbers represent much less mental distance from the traditional view of standards of proof, because b(S) > b(notS) if and only if b(S) > .50. Thus, preponderance could retain a meaning of likelihood exceeding 50%, while clear and convincing means much more likely than 50% and beyond a reasonable doubt means almost certainty. This alternative renders my conceptualization much less jarring, and it also demonstrates that I did not pull my formulations out of thin air.

Yet, I resist taking that normalization route. First, converting to additive beliefs would reintroduce the probabilistic imaging that originally led us astray into all the problems and paradoxes of the traditional view. Second, I contend that directly comparing Bel(S) and Bel(notS) actually conforms better to the actual law than the probabilistic view does. Third, normalization requires measurement of b(S) and b(notS), a step otherwise unnecessary, and a step that is much more difficult for humans than relative judgment.

The evidence at trial will support S to an extent while supporting notS to another extent, and the reformulated standards say that the fact-finder need only compare these two fuzzy beliefs. How does the current law actually state, say, preponderance of the evidence? Consider a couple of classic cases.

In *Livanovitch v. Livanovitch*,[173] the trial court gave the following charge: "If . . . you are more inclined to believe from the evidence that he did so deliver the bonds to the defendant, even though your belief is only *the slightest degree* greater than that he did not, your verdict should be for the plaintiff."[174] The appellate court said:

---

173   131 A. 799 (Vt. 1926).

174   *Id.* at 800.

> The instruction was not erroneous. It was but another way of saying that the slightest preponderance of the evidence in his favor entitled the plaintiff to a verdict. . . . All that is required in a civil case of one who has the burden of proof is that he establish his claim by a preponderance of the evidence. . . . When the equilibrium of proof is destroyed, and the beam inclines toward him who has the burden, however slightly, he has satisfied the requirement of the law, and is entitled to the verdict. "A bare preponderance is sufficient, though the scales drop but a feather's weight." This rule accords with the practice in this state as remembered by the justices of this court, and is well supported by the authorities.[175]

In *Lampe v. Franklin American Trust Co.*,[176] one of the defendant's contentions was that the note in suit had been altered after it had been signed by the defendant's decedent. The trial court refused the defendant's request for an instruction that the jury should find that the instrument was not the decedent's note

> if you find and believe that it is *more probable* that such changes or alterations have been made in the instrument after it was signed by the deceased and without his knowledge and consent, than it is that such alterations and changes were made at or about the time that the deceased signed the instrument and under his direction and with his knowledge and consent.[177]

Holding the refusal to have been proper, the appellate court said:

> The trouble with this statement is that a verdict must be based upon what the jury finds to be facts rather than what they find to be "more probable." . . . This means merely that the party, who has the burden of proof, must produce evidence, tending to show the truth of those facts, "which is more convincing to them as worthy of belief than that which is offered in opposition thereto."[178]

These two cases' formulations sound contradictory. But if one interprets the quotations as speaking in terms of the coarsely gradated belief in the fact compared with the coarsely gradated belief in the fact's negation, based on the evidence presented, the apparent contradiction evaporates. They both seem to be saying that the burdened party should win if and only if Bel(S) > Bel(notS).

Other courts sometimes express more divergent views of the standard of proof. Some writers conclude that courts interpret preponderance in one of three ways: (1) "more convincing," which requires the burdened party to tell a better tale than the opponent tells; (2)

---

175   *Id.*
176   96 S.W.2d 710 (Mo. 1936).
177   *Id.* at 723.
178   *Id.* (quoting Rouchene v. Gamble Constr. Co., 89 S.W.2d 58, 63 (Mo. 1935)).

"more likely than not," which requires a showing of the fact's existence stronger than the showing of its nonexistence; or (3) "really happened," which requires a showing by evidence of what probably transpired outside in the real world.[179] My approach would conform to the middle option of (2), rather than either (1) relative plausibility or (3) absolute measure.

In the end, I submit that comparison of coarsely gradated beliefs is an accurate representation of what the law tells a fact-finder to do with a standard of proof. In civil cases, the fact-finder has to find that Bel(S) is more likely than *not*, which means Bel(S) > Bel(notS). Or as the judge tells the jurors, preponderance means that the evidence "produces in your minds belief that what is sought to be proved is more likely true than not true"[180] or "more probably true than false."[181] By literally instructing fact-finders to decide between S and notS, the law effectively urges them to focus on those two fuzzy beliefs and compare them.

3.   Implications of Reformulation

My views, then, are not seditious. Overall I merely contend, in accordance with the new logic's teaching, that the law charges fact-finders to form a set of fuzzy beliefs, while leaving some belief uncommitted in the face of imperfect evidence, and then to apply the standard of proof by comparing their resultant belief in the burdened party's version to their belief in its negation. Many observers of the legal system would find that contention, putting its slightly new vocabulary to the side, unobjectionable.

Tracing the implications of my contention reveals its hidden powers, however. It implies that the fact-finders at the end of a case would properly apply the standard to each separate element. It also implies that the fact-finders should start the case, being in a state of ignorance

---

179   *See* COVINGTON, *supra* note 113, at 99–100.

180   3 O'MALLEY ET AL., *supra* note 119, § 104.01:

> "Establish by a preponderance of the evidence" means evidence, which as a whole, shows that the fact sought to be proved is more probable than not. In other words, a preponderance of the evidence means such evidence as, when considered and compared with the evidence opposed to it, has more convincing force, and produces in your minds belief that what is sought to be proved is more likely true than not true.

181   Nissho-Iwai Co. v. M/T Stolt Lion, 719 F.2d 34, 38 (2d Cir. 1983) ("The term 'preponderance' means that 'upon all the evidence . . . the facts asserted by the plaintiff are more probably true than false.'" (quoting Porter v. Am. Exp. Lines, Inc., 387 F.2d 409, 411 (3d Cir. 1968))); *see* McBaine, *supra* note 79, at 261–62; Morgan, *supra* note 167, at 66–67.

with lack of proof, at a zero belief. Thus, two paradoxes in the nature of legal proof simply vaporize. The four parts of this Article generate related insights.

First, the linguistic evaluations that humans tend to use in their fuzzy logic, as opposed to quantifications, nicely express the law's development of a coarsely gradated scale of possibilities and probabilities: (1) slightest possibility, (2) reasonable possibility, (3) substantial possibility, (4) equipoise, (5) probability, (6) high probability, and (7) almost certainty. And the coarseness of the scale of likelihood means that the fact-finder in comparing beliefs will not have to draw paper-thin distinctions.

Second, when the fact-finders face multiple elements, it has long appeared that they seek the most believable story by applying the standard of proof to each element. But theorists worry that this conjoined story itself may not meet the standard of proof. Rest assured, because the law knows what it is doing. The MIN operator demonstrates that belief in the conjunction will match the belief in the least likely element, which has already passed the standard of proof.

Third, the notion of burden of proof becomes much clearer. The paradoxical difficulties in applying the burden to weak proof dissipate. For an example, a directed verdict motion by a civil defendant meshes the burden of production with the new view of the preponderance standard. The motion requires the judge to ask if no reasonable jury could view $\text{Bel}(S) > \text{Bel}(\text{notS})$.[182] At the end of the plaintiff's case, if a reasonable $\text{Bel}(\text{notS})$ equals 0 (effectively a "slightest possibility"), then the inequality requires a compatibly reasonable $\text{Bel}(S)$ to exceed 0 (effectively a "reasonable possibility"). That the plaintiff must have established a reasonable possibility is the embodiment of the burden of production, and it is what keeps the plaintiff from surviving with a mere feather's weight of evidence. An illustrative situation would be where the plaintiff has produced a little evidence, but it

---

182   The reference to a "reasonable" jury reflects the fact that on such a motion the judge is reviewing the jury's hypothesized application of the standard of proof. The judge's standard of review turns on whether a jury could not reasonably, or rationally, find for the non-movant. That is, the defendant must show that a verdict for the plaintiff, given the standard of proof, is not reasonably possible. *See* Clermont, *supra* note 20, at 1126–27.

We can state this standard of *review* simply and fuzzily in terms of the law's coarsely gradated scale of possibilities and probabilities, without the complications that belief functions impose on the standard of *proof*. The reason is that we do not expect the judge to retain uncommitted belief in applying a standard of review. The "evidence" for applying the standard is complete. We want from the judge the likelihood of jury error in finding for the plaintiff, with the complement being the likelihood of jury correctness in finding for the plaintiff.

is "pure" evidence that gives the defendant no support.[183] If a reasonable jury could find for the plaintiff on such proof, the judge should deny the directed verdict motion. If the defendant then produces no effective evidence during the rest of the trial, but moves again for a directed verdict at the end of all the evidence, the judge should deny the motion and the case should go to the jury. The jury, if it were to take the same view of the evidence as the judge hypothesized, could find for the plaintiff—even on such thin evidence.

Fourth, a new understanding of how to apply the standard of proof to the party with the burden of persuasion follows naturally, even if not inevitably, from the foregoing logical conceptualization of the nature of proof. The standard should concede that upon incomplete, inconclusive, ambiguous, dissonant, and untrustworthy proof, some of our belief will remain indeterminate. The standard should look only to committed belief, comparing belief in the burdened party's version versus disbelief.

Not only does this comparative approach comport with the natural cognitive method that follows from telling the fact-finders they must decide for one side or the other, but also it does nothing to interfere with the current procedural and substantive functioning of the standard of proof. For example, the traditional view of the preponderance standard as a showing of a probability greater than 50% appeared appropriate for civil cases: among competing fixed standards,[184] it minimizes the expected number of erroneous decisions

---

183   *See* Liu & Yager, *supra* note 126, at 18–19 (discussing Liebniz's notions of pure and mixed evidence).

184   If we knew more about the base rates for the type of case or the realities of the particular case itself, we might want to adjust the standard of proof. For example, a variable standard of proof, set on a case-by-case basis by the ideal judge, could serve accuracy by offsetting the unavailability or inadmissibility of evidence in the particular case. *See* Dominique Demougin & Claude Fluet, *Deterrence Versus Judicial Error: A Comparative View of Standards of Proof*, 161 J. INSTITUTIONAL & THEORETICAL ECON. 193 (2005) (arguing by sophisticated analysis for a variable standard of proof). More generally, in an idealized system, one could argue that the standard of proof should slightly vary issue-by-issue in response to the expected utility of each outcome. *See* Richard A. Posner, *An Economic Approach to Legal Procedure and Judicial Administration*, 2 J. LEGAL STUD. 399, 414–16 (1973) (using economic analysis); *cf.* Richard D. Friedman, *Standards of Persuasion and the Distinction Between Fact and Law*, 86 Nw. U. L. REV. 916, 926 (1992) (extending the arguments to law-determining). But the path of the law has not been toward variable standards of proof but instead toward standards generally applicable for whole categories of cases—while making gross adjustments as to whole categories of issues when substantive considerations, such as the high social cost of criminally convicting the innocent, counsel adjustment.

and also the expected sum of wrongful amounts of damages,[185] which is the goal that the law apparently pursues in preference to optimizing incentives for primary conduct.[186] My reformulated standard has the same error-minimizing properties, but achieves them in the real world where the law of the excluded middle does not hold and where some indeterminacy prevails. For an idea of a proof adapted from a probabilist's proof, let $b(S)=p$ be the apparent probability that the defendant is liable (for D dollars) under a two-valued view. If Bel(S) > Bel(notS), then $p > \frac{1}{2}$; call p by the name p_1 in that case. If Bel(S) ≤ Bel(notS), call it p_2. On the one hand, under the preponderance

---

185   This argument for the preponderance standard is strong, because it seems demonstrably optimal given two conditions that are plausible. The first condition is that an error in favor of the plaintiff is neither more undesirable nor less undesirable than an error in favor of the defendant, or that a dollar mistakenly paid by the defendant (a false positive) is just as costly to society as a dollar mistakenly uncompensated to the plaintiff (a false negative). The second condition is that the goal is to minimize the sum of expected costs from these two types of error, that is, the system wants to keep the amounts suffered mistakenly to a minimum.

    Accepting that these conditions generally prevail outside the criminal law—and discounting more intangible possibilities, such as there being differential perceptions of loss and gain or varying marginal utilities of wealth that are worthy of consideration—the preponderance standard should perform better than any other non-variable standard of proof. The reason is that by so deciding in accordance with apparent probabilities, the legal system in the long run will make fewer errors than, for example, the many false negatives that a virtual-certainty standard would impose. The preponderance standard also minimizes the system's expected error costs. Let p be the apparent probability that the defendant is liable (for D dollars). If $p > \frac{1}{2}$, call it p_1; and if $p \leq \frac{1}{2}$, call it p_2. On the one hand, under the preponderance standard, the expected sum of false positives and false negatives over the run of cases is $\Sigma[(1-p_1)D + p_2D]$. On the other hand, under a very high standard of proof that eliminates false positives, the analogous sum is $\Sigma[p_1D + p_2D]$. Given that $(1-p_1)$ is less than p_1, the preponderance standard therefore lowers the system's expected error costs. *See* D.H. Kaye, *The Error of Equal Error Rates*, 1 LAW PROBABILITY & RISK 3, 7 (2002) ("The general appeal of the $[p > \frac{1}{2}]$ rule lies in the fact that it minimizes expected losses." (citation omitted)); David Hamer, *Probabilistic Standards of Proof, Their Complements and the Errors That Are Expected to Flow from Them*, 1 U. NEW ENG. L.J. 71 (2004); *cf.* Neil Orloff & Jery Stedinger, *A Framework for Evaluating the Preponderance–of–the–Evidence Standard*, 131 U. PA. L. REV. 1159 (1983) (considering bias in the distribution of errors).

186   It could well be that the goal should not focus only on minimizing the sum of expected costs from the two types of erroneous decisions measured from an ex post perspective. Going forward from decision, correct decisions matter too, in that they increase the deterrent effect and reduce the chilling effect of the law's applications. From a social welfare point of view, the law should set the standard of proof only after taking these effects into account. *See* Louis Kaplow, *Burden of Proof*, 121 YALE L.J. 738 (2012); *see also* Fredrick E. Vars, *Toward a General Theory of Standards of Proof*, 60 CATH. U. L. REV. 1 (2010) (using a sophisticated utility analysis to set the standard for the issue of mental incapacity in will contests).

standard, the expected sum of false positives and false negatives over the run of cases is $\Sigma[(1-p_1)D + p_2D]$. On the other hand, under a very high standard that eliminates false positives, the analogous sum is $\Sigma[p_1D + p_2D]$. Therefore, given that $(1-p_1)$ is less than p_1, the reformulated preponderance standard lowers the system's expected error costs.

To close, a comprehensive example would perhaps be beneficial. Suppose that someone has seriously injured Suzie, in circumstances suggesting fault. She sues Tom, which means that she must prove his identity as the tortfeasor—as well as fault, causation, and injury. She introduces a fair amount of evidence.

First, the fact-finder would assess that evidence and might conclude as follows: (1) The evidence points to Tom being the perpetrator. If the fact-finder were a bettor, he would put the odds at 3:2, or 60%. Using words, he would say that Tom was probably the perpetrator. (2) The question of fault was a tough one. There are uncertainties as to what was done, but there is also a vagueness concerning how blameworthy the supposed acts really were. The fact-finder needs commensurable measures, so that he can evaluate a mix of random and nonrandom uncertainty. If forced to assess all the evidence on this issue and put it on a scale of truth running from zero to one, he would say .7. He might feel more comfortable saying fault was probable. (3) The acts, whatever they were, apparently caused the injury. Proximate cause is about as vague and multivalent as a legal concept can get. The fact-finder is pretty convinced nevertheless. He would put causation at .8, or highly probable. (4) Suzie's injuries are not really very vague or uncertain. He would put this element of the tort at .95, or beyond a reasonable doubt. Note that the new conceptualization changes nothing, to this point, regarding the fact-finder's task as traditionally envisaged.

Second, the fact-finder may want to combine these findings. They are a mixture of probabilities and degrees of truth. But viewing them all as degrees of truth invokes the MIN operator, so that he can say that Suzie's story comes in at .6, or probable. Suzie should win, by the use of fuzzy logic.

Third, this approach does not do a terribly good job of accounting for the state of the evidence. It still poses an odd question to the fact-finder: given imperfect evidence, what is the degree to which the plaintiff is right? Belief functions work better here to reflect the fact-finder's actual knowledge: belief starts at zero, and some belief will remain uncommitted in the absence of perfect evidence. That is, on a fact to which the standard of proof applies, the belief function route is the one to take, rather than invoking the simplistic scale of likelihood

just described. Instead of saying that Tom's fault is probable, the fact-finder should speak and think in terms of degrees of belief.

Fourth, although belief functions do not require placement on a scale, the fact-finder in effect might end in believing Suzie's position on Tom's fault to be only substantially possible. That situation does not mean that Suzie should lose, however. The fact-finder might, if forced to express likelihood, believe the falsity of Tom's fault merely to a reasonable possibility. All the fact-finder must do is to compare belief and disbelief: all that preponderance of the evidence requires is that the strength of the fact-finder's belief that Tom was at fault must exceed his belief that Tom was not at fault. Belief functions thus add the idea that the fact-finder in such a case must have a belief in the case's truth stronger than his belief in its falsity. While some of the fact-finder's belief remains uncommitted, he did find Suzie's position to be a good one: more likely true than false. So, Suzie should still win, by the use of belief functions.

Conclusion

This Article deploys the new logic—in particular, fuzzy logic and belief functions in their broad senses—to conceptualize the standards of proof. This was not a heavily prescriptive endeavor, which would have tried to argue normatively for the best way to apply standards. Instead, it was mainly a descriptive and explanatory endeavor, trying to unearth how standards of proof actually work in the law world. Compared to the traditionally probabilistic account, this conceptualization conforms more closely to what we know of people's cognition, captures better what the law says its standards are and how it manipulates them, and improves our mental image of the fact-finders' task. One virtue of the conceptualization is that it is not radically new, as it principally acts to confirm the law's ancient message that fact-finders should simply compare their non-quantified views of the fact's truth and falsity. The conceptualization leaves the law's standards essentially intact to accomplish their current purposes. Another virtue is that it nevertheless manages to resolve some stubborn problems of proof, including the fabled conjunction paradox. Thus, for understanding the standards of proof, degrees of fuzzy belief work better than traditional probabilities.

In brief, the new logic reveals that the law wants fact-finders to form degrees of belief that would conform to a fuzzy scale, to combine them in a logical fashion while leaving some belief uncommitted in the face of imperfect evidence, and then to apply the standard of proof by comparing their resultant belief in the burdened party's version of fact to their belief in its negation.