Cornell Law Library
Scholarship@Cornell Law: A Digital Repository

Cornell Law Faculty Working Papers | Faculty Scholarship

2-4-2012

Death of Paradox: The Killer Logic Beneath the Standards of Proof

Kevin M. Clermont
Cornell Law School, kmc12@cornell.edu

Follow this and additional works at: http://scholarship.law.cornell.edu/clsops_papers

Part of the Civil Procedure Commons

Recommended Citation
Clermont, Kevin M., "Death of Paradox: The Killer Logic Beneath the Standards of Proof" (2012). *Cornell Law Faculty Working Papers.* Paper 94.
http://scholarship.law.cornell.edu/clsops_papers/94

This Article is brought to you for free and open access by the Faculty Scholarship at Scholarship@Cornell Law: A Digital Repository. It has been accepted for inclusion in Cornell Law Faculty Working Papers by an authorized administrator of Scholarship@Cornell Law: A Digital Repository. For more information, please contact jmp8@cornell.edu.

Feb. 4, 2012

Death of Paradox:
The Killer Logic Beneath the Standards of Proof

Kevin M. Clermont*

The prevailing but contested view of proof standards is that factfinders should determine facts by probabilistic reasoning. Given imperfect evidence, they should ask themselves what they think the chances are that the burdened party would be right if the truth were to become known; they then 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 probability. This modern logic suggests that factfinders view evidence of an imprecisely perceived and described reality to form a fuzzy degree of belief in a fact's existence; they then apply the standard of proof in accordance with the theory of belief functions, by comparing their belief in a fact's existence to their belief in its negation.

This understanding explains how the standard of proof actually works in the law world. It gives a superior mental image of the factfinders' task, conforms more closely to what we know of people's cognition, and captures better what the law says its standards are and how it manipulates them. One virtue of this conceptualization is that it is not a radically new view. Another virtue is that it nevertheless manages to resolve some stubborn problems of proof, including the infamous conjunction paradox.

Introduction.. 2

I. Assessing Evidence. 9
A. Theories. 9
1. Psychology Theories.. 9
2. Probability Theories. 11
3. Zadeh's Fuzzy Logic. 12
B. Legal Application: Gradated Likelihood. 21

II. Combining Assessments. 23
A. Fuzzy Operators. 23
1. Maximum and Minimum. 23
2. Product Rule Contrasted.. 24
3. Negation Operator. 27
B. Legal Application: Conjunction Paradox. 28

III. Analyzing Beliefs. 32
A. Shafer's Belief Functions. 32
1. Basics of Theory.. 32
2. Negation Operator. 34
3. Lack of Proof. 34
B. Legal Application: Burden of Production.. 35

IV. Applying Standards. 38
A. Comparison of Beliefs.. 38
B. Legal Application: Burden of Persuasion.. 39
1. Traditional View. 40
2. Reformulated View. 41
3. Implications of Reformulation. 47

Conclusion. 51

---

* Ziff Professor of Law, Cornell University. I want to thank my philosophical wife, Emily Sherwin, for getting me to think about vagueness! Then came some incredibly helpful comments from Ron Allen, Adrienne Clermont, Simona Grossi, Robert Hockett, John Leubsdorf, and John Palmer, even if a few of the commenters (but not my family members) disagreed sharply with my position.

Electronic copy available at: http://ssrn.com/abstract=1986346

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 . . . .[**]

INTRODUCTION

The wonderful "new evidence" scholarship has made tremendous strides 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] Although new, this work represented the necessary return to abandoned efforts by past greats such as Wigmore.[2] But the new workers tended to remain too wedded, or overly hostile, to subjective probabilities for evaluating evidence[3] and Bayes' theorem for combining evidence,[4] and so caused the debates to become "unproductive and

---

[**]Marcel Proust, Remembrance of Things Past, vol. V, ch. II (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 (2007) (similar).

[1]See generally William Twining, Rethinking Evidence 237-48 (2d ed. 2006); Richard Lempert, The New Evidence Scholarship, in Probability and Inference in the Law of Evidence: The Uses and Limits of Bayesianism 61 (Peter Tillers & Eric D. Green eds., 1988).

[2]See, e.g., Terence Anderson, David Schum & William Twining, Analysis of Evidence (2d ed. 2005) (building on John Henry Wigmore, The Science of Judicial Proof As Given by Logic, Psychology, and General Experience, and Illustrated in Judicial Trials (3d ed. 1937)).

[3]See generally Richard Eggleston, Evidence, Proof and Probability (2d ed. 1983); Symposium, Decision and Inference in Litigation, 13 Cardozo L. Rev. 247 (1991); Symposium, Probability and Inference in the Law of Evidence, 66 B.U. L. Rev. 377 (1986); cf. Vaughn C. Ball, The Moment of Truth: Probability Theory and Standards of Proof, 14 Vand. L. Rev. 807, 809-12 (1961) (treating frequency 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).

[4]See generally Sharon Bertsch McGrayne, The Theory That Would Not Die (2011) (recounting the centuries of controversy generated by Bayes' theorem).

Electronic copy available at: http://ssrn.com/abstract=1986346

sterile."[5] In any event, the debates left unsolved lots of troubling problems and paradoxes in our law of evidence.

The "new logic."—One specific diagnosis of this shortcoming is that the new evidence tended to neglect the contemporaneous advances in logic.[6] The new, so-called nonclassical logic looks and sounds much like standard logic but refuses to accept some critical assumption.[7] Most often, the assumption rejected is that every proposition must either be true or be false, an assumption called the law of bivalence. But if one entertains the classically excluded middle so that both P and notP can be true and false to a degree, then one can show that sometimes P can equal notP, which is a rather disquieting contradiction.[8] Thus, fashioning the new logic faced some challenges.

The first elaboration of the new logic of special interest to lawyers relates to and builds on the modern philosophy, beginning with Bertrand Russell's work, that struggled with the problem of vagueness.[9] Work on vagueness addresses matters such as the famed sorites paradox of ancient Greece ("sorites" comes from the Greek word for heap):

- 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

---

[5]Peter Tillers, Trial by Mathematics—Reconsidered, 10 Law Probability & Risk 167, 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 Comment. on Evid., iss. 1, http://www.bepress.com/ice/vol8/iss1/art1 (2010).

[6]See Tillers, supra note 5, at 171 (suggesting that diagnosis). 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).

[7]See Theodore Sider, Logic for Philosophy 72-73 (2010).

[8]See Peter Suber, Non-Contradiction and Excluded Middle, http://www.earlham.edu/~peters/courses/logsys/pnc-pem.htm.

[9]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, Vagueness (1996); Hartry Field, No Fact of the Matter, 81 Australasian J. Phil. 457 (2003) (countering the Williamson view); Hartry Field, Vagueness, Partial Belief, and Logic, in Meanings and Other Things (G. Ostertag ed., 2012), available at http://philosophy.fas.nyu.edu/docs/IO/1158/schiffer2004b.pdf (incorporating ideas similar to fuzzy logic); Hartry Field, Indeterminacy, Degree of Belief, and Excluded Middle, 34 Noûs 1, 20 (2000) (referencing work on belief functions).

Electronic copy available at: http://ssrn.com/abstract=1986346

> you will by logic still have a heap. But there is no heap. Thus, heap equals non-heap. 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 non-heap. But was there a fixed boundary? No!—at least according to most. A different path taken in the attempt to avoid the paradox leads to the embrace of *many-valued logic*.[10] This form of logic boldly declines the simplification offered by two-valued 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 non-heap, but something that falls in between, with the statement "this is a heap" being both true and not true.[11]

The second interesting elaboration of the new logic involves developments in the field of *imprecise probability*.[12] This field provides a useful extension of probability theory whenever information is conflicting or scarce. 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, classical probability (by which I mean the doctrine of random chance as well as subjective probability) appears as a special case in this theory. The rules associated with classical probability carry over, except those based on assuming an excluded middle.

All this logic may be new, but it has an extended history of forerunners. Problems of many-valued logic have troubled thinkers since before Aristotle,[13] even if their thoughts found more

---

[10]See generally J.C. Beall & Bas C. van Fraassen, Possibilities and Paradox: An Introduction to Modal and Many-Valued Logic (2003).

[11]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: The New Science of Fuzzy Logic 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., 1999) (saying that 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 (Petr Cintula et al. eds., forthcoming 2012), available at http://www-personal.usyd.edu.au/~njjsmith/papers/SmithFuzLogHOVag.pdf (similar). See generally Liars and Heaps (J.C. Beall ed., 2004); 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).

[12]See generally Peter Walley, Statistical Reasoning with Imprecise Probabilities (1991).

[13]See James F. Brulé, Fuzzy Systems—A Tutorial, http://www.austinlinks.com/Fuzzy/tutorial.html (1998) (recounting the history briefly).

receptive soil in the East than in the West.[14] Imprecise probability goes back to the nineteenth century.[15] Nevertheless, the new logic has enjoyed a recent flowering, inspired by the development of quantum mechanics and instructed by those advances in philosophy and probability.

"Fuzzy logic."—The particular bloom known as fuzzy logic finds its roots in the seminal 1965 article by Lotfi Zadeh.[16] He used degrees of membership in a *fuzzy* set running from 1 to 0, rather than 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.[17] It is instead a formal system of logic, one that is by now highly developed and hence rather complicated.[18] Of course, it has become so well-known and dominant because of its countless practical applications, especially in the computer business and consumer electronics.[19] 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. 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 understandings of likelihood.[21] But it is not an exclusive tool. "In order to treat different aspects of the same problems,

---

[14]See Kosko, supra note 11, at 69-78 (giving an impassioned attack on the West's hostility to many-valued logic, but also fleshing out the historical account).

[15]See George Boole, An Investigation of the Laws of Thought on Which Are Founded the Mathematical Theories of Logic and Probabilities (London, Walton and Maberly 1854) (doing the early work).

[16]L.A. Zadeh, Fuzzy Sets, 8 Info. & Control 338 (1965).

[17]Zadeh substituted the term "fuzzy" for "vague" more to be provocative than to be descriptive. See Kosko, supra note 11, at 19-20, 142, 145, 148.

[18]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).

[19]See Kosko, supra note 11, at 157-200 (describing fuzzy computer systems).

[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 6, 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, e.g., Richard M. Royall, Statistical Evidence: A Likelihood Paradigm 5-6, 28 (1997). But for most people, likelihood means probability. I use it 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."

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 here is Glenn Shafer's imposing elaboration of imprecise probability from 1976.[23] His work on belief functions built a bridge between fuzzy logic and classical probability, and in the process nicely captured our legal decisionmaking scheme.[24] Beliefs, again on a scale running from 1 to 0, can express uncertainty about the evidence by allowing for representation of ignorance and indeterminacy.[25]

Belief functions manage this task without being a mere scheme of second-order probability,[26] which has failed both the statistically minded[27] and psychologically minded[28] as a way to explain

---

[22]Mircea Reghiş & 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 Evidential Reasoning 41, 200-01 (1994) (disbelieving that it is "possible to capture all of this behavioral richness within the confines of any single formal system of probabilities").

[23]Glenn Shafer, A Mathematical Theory of Evidence (1976) (using "evidence" in a much broader sense than legal evidence), reviewed by Lotfi A. Zadeh, 5 AI Mag., no. 3, 1984, at 81, 83, available at www.aaai.org/ojs/index.php/aimagazine/article/download/452/388 (treating Shafer's theory as a version of fuzzy logic's possibility theory); Glenn Shafer, Perspectives on the Theory and Practice of Belief Functions, 4 Int'l J. Approximate Reasoning 323 (1990), available at http://www.personal.psu.edu/students/r/j/rjc157/shafer_1990.pdf.

[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]For an accessible introduction, see Schum, supra note 22, at 222-43.

[26]See Schum, supra note 6, at 868 (distinguishing belief functions from second-order probability).

[27]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). That article advanced a new theory of standards of proof based on the statistical concept of confidence intervals. It described how sure a factfinder 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) (conceding that confidence relates solely to the probability of avoiding false positives), with D.H. Kaye, Commentary,

standards of proof. Nor are belief functions a bow to some anti-probabilist yearning to go with the factfinders' feelings. Instead, they rest on a highly rigorous mathematical base. And they use "belief" to invoke not some squishy personal view but rather the factfinders' attempt objectively to express their degree of certainty about the state of the real world as represented by the evidence put before them.[29]

Relation to law.—My thesis is that a better explanation of what the law does as to proof lies in the new logic than in classical 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. Aristotlean two-valued logic did not work for such partial truths. Law seemed, early and intuitively, to draw many-valued logic from natural language and daily life. But in the late eighteenth century, law and nascent probability theory began interacting.[30] The law became more open about accepting uncertainty, but otherwise probabilistic theorizing only lightly influenced the law until it began quite

---

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).

[28]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 factfinder 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 factfinder's world knowledge, and complete without striking gaps in expected components), and uniqueness (which means the absence of plausible alternative stories). The factfinder achieves this by an automatic or unconscious process. The clearer the view of the case that the chosen story delivers to the factfinder, the more confident the factfinder will be. Against this level of confidence, the factfinder 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).

[29]See Glenn Shafer, The Construction of Probability Arguments, 66 B.U. L. Rev. 799, 801-04 (1986). But cf. 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).

[30]See Kevin M. Clermont & Emily Sherwin, A Comparative View of Standards of Proof, 50 Am. J. Comp. L. 243, 256-57 (2002).

recently a major assault.[31] Still, classical 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. 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. The prevailing but contested view is that factfinders should determine facts by 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 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 factfinder's complexly constructed belief is the more relevant question. We are not as concerned with how certain the factfinder is in a world of random uncertainty, as we are with the degree of truth the factfinder constructs in a world of vague imprecision. The next step builds on the idea that factfinder 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 evidence, some of the factfinder's belief should remain indeterminate. The output is not a probability, but what a logician would call a nonadditive degree of belief.[32] In the final step, the standard of proof will call on the factfinders to weigh that degree 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 probability theory, and they better capture the effect of imperfect evidence. Also, abandoning probabilities opens the door to the logicians' 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 the 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 view of proof works so well in this regard that it appears to have been all along the law's intuitive conceptualization of proof.

---

[31] Arguably, the first effective volleys of that assault on the law 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).

[32] See Shapira, supra note 24, at 1613-16 (distinguishing additive from nonadditive). Consequently, a belief and a belief in its negation will often not add to one.

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 factfinding: (I) assessing evidence on related points, (II) combining separate assessments of independent points, (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 a key feature of the law of proof: (1) the gradated scale of likelihood, (2) the conjunction paradox, (3) the burden of production, and (4) the burden of persuasion.

The primary aim of the article is descriptive and explanatory, not prescriptive other than by implication. I seek to reveal what the law actually tells its factfinders to do. Over the centuries the law's charge to factfinders has evolved, by a process entailing considerations of both ideal factfinding and also human limitations. But I am not championing new ideals or new limitations. My real interest here lies in elaborating the proof process that the law has chosen.

As to that elaboration, I believe and shall try to demonstrate that the law embraces the new logic. That version of logic captures the epistemic function of law better than classical probability. A better 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 factfinders actually decide. It could lead to legal reforms, such as clearer instructions on the standards of proof. But that is not the present project. Indeed, the message of this article is that the law currently does a much better job in structuring factfinding than one would guess by reading its many critical commentators.

I. Assessing Evidence

This Part will convey the basics of fuzzy logic 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 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 factfinder 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.[33]

---

[33]See Clermont, supra note 28, 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.[34] 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 a 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*.[35] 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 decisionmaker 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 decisionmaker would then combine these into a weighted average that determines the fact's existence.

Even if that theory reflects reality, it is clear that humans do not naturally use perfectly "rational" techniques, but instead use less accurate "intuitive" techniques.[36] Those techniques are subject to all sorts of heuristics and other biases.[37] Any account would also have to incorporate the dual processes, automatic and systematic, of cognitive processing.[38]

---

[34]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).

[35]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 93.

[36]See generally Daniel Kahneman, Thinking, Fast and Slow (2011). For treatment of humans' bounded rationality, see infra text accompanying note 79, 80.

[37]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).

[38]See Anna Ronkainen, Dual-Process Cognition and Legal Reasoning, http://ssrn.com/abstract=1879429, at 1 (June 30, 2011) ("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.").

In actuality, then, the factfinders' performance of the cognitive process would usually be approximate and nonquantitative. The factfinders 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 factfinders can do, actually do, or think they are doing, on the one hand, is what the law should or does tell the factfinder to do, on the other hand. Surprisingly, even mathematicians cannot agree on how a factfinder should perform the task of processing evidence, and so cannot unanimously guide the law on an ideal path to take.[39] The most popular candidate for rigorously evaluating and combining related evidence is the Bayesian approach, built on Pascalian probabilities.

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 play a broad role in legal factfinding.[40] 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 factfinder to begin.[41] 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.[42]

---

[39]See, e.g., Symposium, Artificial Intelligence and Judicial Proof, 22 Cardozo L. Rev. 1365 (2001).

[40]See, e.g., Pennington & Hastie, supra note 35, at 262-68; Shafer, supra note 29, 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).

[41]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).

[42]See Lea Brilmayer & Lewis Kornhauser, Review: Quantitative Methods and Legal Decisions, 46 U. Chi. L. Rev. 116, 135-48 (1978).

Third, its mathematical approach is not realistic, of course. It does not conform to the way humans arrive at prior probabilities or the way they combine them with new evidence to produce posterior probabilities.[43]

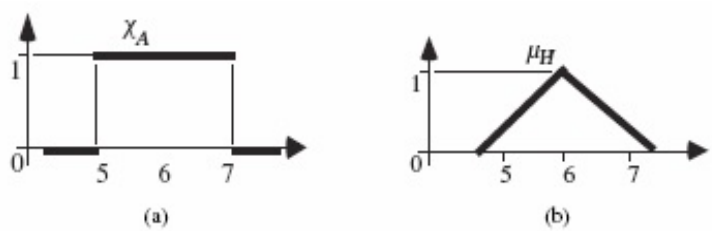
3. Zadeh's Fuzzy Logic

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 in the universe X may take values throughout the whole interval [0,1], rather than just the two values of 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, or $5.0' \leq x \leq 7.0'$. 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.

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. A 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) representing a fuzzy set of men somewhere near six feet tall:[44]



b. Linguistics.—Important to note is the role of qualifying language in the prior example. *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 are not 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 then

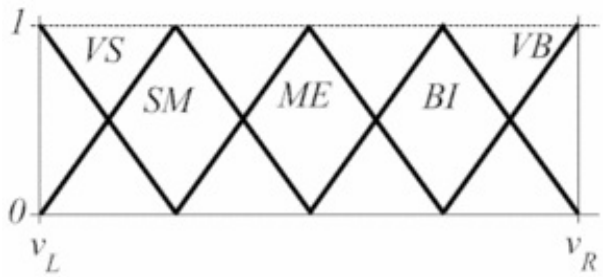
---

[43]Compare Samuel Kotz & Donna F. Stroup, Educated Guessing: How to Cope in an Uncertain World (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 used in law).

[44]The figure comes from Ross & Parkinson, supra note 18, at 30.

use them for classification, decisionmaking, 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 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:[45]



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 triangles at the top, perhaps equating to very small, almost medium, almost big, and very big. (The representation might be more accurate if the apex of each of the four triangles at the bottom extended to a value of 1.) 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 standard logic with its excluded middle. In the real world, boundaries of things often appear indistinct, making important, even essential, the human ability to process fuzzy information.

Indeed, some research suggests that such verbal categories work better than numerical translations.[46] Other research indicates that they work best as a matter of natural language if the categories are equally sized and shaped.[47] It might even be feasible to construct and employ an

---

[45]The figure comes from Vilém Novák, Modeling with Words, http://www.scholarpedia.org/article/Modeling_with_words (2008).

[46]See Alf C. Zimmer, Verbal vs. Numerical Processing of Subjective Probabilities, in Decision Making Under Uncertainty 159, 180 (Roland W. Scholz ed., 1983); cf. Clermont, supra note 20, at 1144-50 (discussing the verbalization of legal standards more generally, and recounting the empirical studies that show factfinders' difficulties in comprehending standards of proof); 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 fuzzy logic, and highlighting "the need to reevaluate the reliability and validity of past research findings on quantifying" standards of proof).

[47]See Zimmer, supra note 46, at 166.

ordinally ranked vocabulary of likelihood that allows effective interpersonal communication.[48]

c. Evaluating and combining evidence.—Fuzzy logic has formal rules for combining related evidence,[49] once evaluated by one method or another.[50] Moreover, to jump ahead to the belief functions that will be discussed later, the theoretical work thereon consists mainly of developing tools for combining evidence. In particular, its very prominent Dempster-Shafer rule governs the combination of assessments.[51] That rule is very complicated, because it abstractly addresses the problem in the most general terms possible (Bayes' theorem turns out to be a special case of that approach).[52] The rule is also quite contested, especially because it does not deal well with conflicts in the evidence, and consequently it has generated many competitors.[53]

In the end, these formal approaches are clearly not a realistic representation of human

---

[48]See Mandeep K. Dhami & Thomas S. Wallsten, Interpersonal Comparison of Subjective Probabilities: Toward Translating Linguistic Probabilities, 33 Memory & Cognition 1057 (2005).

[49]See Petr Hajek, Fuzzy Logic, in The Stanford Encyclopedia of Philosophy (Edward N. Zalta ed., Fall 2010), available at http://plato.stanford.edu/archives/fall2010/entries/logic-fuzzy/.

[50]Cf. Brulé, supra note 13 ("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.").

[51]See Glenn Shafer, A Mathematical Theory of Evidence 6, 57-67 (1976) (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 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, at 25.

[52]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 51, at 345.

[53]See Kari Sentz & Scott Ferson, Combination of Evidence in Dempster-Shafer Theory 17-27 (Sandia Nat'l Labs. 2002) (describing thirteen alternatives), available at http://www.sandia.gov/epistemic/Reports/SAND2002-0835.pdf.

cognition.[54] 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 nonquantitative and approximate fashion. Consequently, the law has generally left the assessment of related evidence to its factfinders' instinctive treatment.

Fortunately, fuzzy logic is not dogmatic on the method used to evaluate and combine related evidence. It is compatible with the factfinders' assessing all the related evidence by any means. Just one possibility would be information integration's weighted average, described above.[55] In fact, a weighted-average approach commonly appears in the decisionmaking process of today's fuzzy computer programs.[56]

I therefore will not pursue those formal rules, which constitute the new logic in a narrow sense.[57] My interest in this article is not so much the initial assessment 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 and then to process those views according to established rules of reasoning. That is, fuzzy logic in this broad sense gives a new and effective way to explore humans' expression of their assessment. And it then opens the door to subjecting those expressions to so-called approximate reasoning, which despite its name can be quite rigorously performed.[58]

d. Fuzziness versus probability.—I propose considering the broad version of fuzzy logic as a model of human expression and reasoning in law, in preference to a probability-based view. The significance of adopting it for this purpose becomes evident upon distinguishing a fuzzy statement from a probabilistic one.[59]

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

---

[54]See Schum, supra note 6, at 852-53 (citing the lack of empirical support).

[55]See supra text accompanying note 35.

[56]See Kosko, supra note 11, at 176-80 (describing fuzzy computer systems).

[57]See Novák, supra note 45 ("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.").

[58]See Ronald R. Yager, New Paradigms for Reasoning with Uncertain Information, 13 Cardozo L. Rev. 1005, 1017-24 (1991) (explaining approximate reasoning).

[59]See generally Mark Colyvan, Is Probability the Only Coherent Approach to Uncertainty?, 28 Risk Analysis 645 (2008); Bart Kosko, Fuzziness vs. Probability, 17 Int'l J. Gen. Sys. 211 (1990).

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 measures the chance that the event occurs.

The uncertainty of whether an event occurs or not is normally random, as contrasted with the nonrandom uncertainty of vagueness that fuzziness measures. Probability, with its theory of randomness in a bivalent world, expresses the chance of whether something will occur, all the while knowing it will occur on an all-or-nothing basis. Fuzziness expresses imprecision as a degree of membership, rather than just that random uncertainty.

Fuzziness conveys all the information we have about an event, which most often ends up as a partial truth. Probability conveys what we know, with only partial knowledge, about whether an event occurs. With more information, the uncertainty in the probability will dissipate, and if the event's occurrence becomes known the probability itself will morph to a value of 1 or 0. By contrast, 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's view of the world tends to fuzz up.

Probabilism and fuzziness can describe much more than these 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 people 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 dealing 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 systems.—Which system 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 adequate. 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."[60] The two systems are not interchangeable, even though people tend to treat them so. They 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 the law has made. But at this point for the reader, it is natural to jump to the conclusion that the law in its factfinding 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

---

[60]Kosko, supra note 11, at 15.

perpetrator. Just to slow you up, however, I point out that in legal factfinding 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?[61] 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.[62] 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 litigation.[63]

Weighing pros and cons.—Bearing on the choice, one advantage of fuzzy logic is that it is the more inclusive system. Being a form of many-valued logic, fuzzy logic does not require 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 only sets with strict membership classifications, an in-or-out characteristic symbolized respectively by values of either 1 or 0. But many-valued logic includes two-valued logic. Bivalence is a special, extreme case of multivalent logic. The world of black or white is a special, extreme case of the world shaded in grays.

Another advantage of fuzzy logic is that 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 expresses 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. It becomes particularly troublesome when trying to gauge a probability resting on imperfect evidence that leaves a lot indeterminate. Fuzzy expression of a belief somewhat better handles incomplete, inconclusive, ambiguous, dissonant, and untrustworthy evidence.

In many situations, a fact is subject both to vagueness and to random 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 complex situations.[64] For example, it might be that there was a .70

---

[61]See id. at 44-46 (using this image).

[62]See supra note 9.

[63]See Leo Katz, Why the Law Is So Perverse 139-56 (2011) (cataloging examples, without invoking fuzzy logic). 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 (2008) (listing examples).

[64]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 *ultrafuzzy* set

blameworthy act. Or while the act was completely blameworthy, there was a 70% chance that it occurred. But what about a 70% chance of a .70 degree of fault? Then, random and nonrandom uncertainty needs to be combined. To combine them, a common currency is necessary.

Moreover, a decision may rest on a number of facts, some of which demonstrate only random uncertainty, while others show some vagueness as well. The need in decisionmaking to combine these factfindings counsels use of one logical system, because "we cannot coherently countenance two different kinds of degree of belief."[65]

Fuzzy logic is the more inclusive logical system, suggesting its use for each fact and for combining evaluations of different facts. Indeed, it can effortlessly express even a classical probability as membership in a set, that is, probability is the degree to which the imagined universe of all trials would be a subset of successful trials.[66] This compatibility allows easy combination of a randomly uncertain set with an imprecise set, in accordance with fuzzy logical operators.[67] Fuzzy logic provides

---

or type-2 fuzzy set, operating as a first step toward type-*n* fuzzy sets and in contrast to the type-1 fuzzy sets discussed up to here. See, e.g., Mark Jablonowski, An Ultra-fuzzy Model of Aggregate Growth in Catastrophic Risk Potentials, 2008 Annual Meeting of the North American Fuzzy Information Processing Society, available at http://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4531214. 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., June 2009, available at http://sipi.usc.edu/~mendel/publications/MendelSMCeNewsletter6-09.pdf.

[65]Nicholas J.J. Smith, Degree of Belief Is Expected Truth Value, in Cuts and Clouds: Vagueness, Its Nature and Its Logic 491, 491 (Richard Dietz & Sebastiano Moruzzi eds., 2010), available at http://www-personal.usyd.edu.au/~njjsmith/papers/smith-degrees-belief-truth.pdf (cataloging the difficulties that would come from entertaining both probabilities and degrees of belief, when the two logical systems employ different operators).

[66]See Kosko, supra note 11, at 55-64 (calling this set the whole in the part).

[67]See infra Part II-A (explaining the MIN operator). This approach would handle the uncertainty of whether an imprecise event occurred at all. See, e.g., 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 to account for subjective probabilities at the same time).

the needed common currency.

In sum, probability has the advantage of bivalent simplicity, but it will often be inaccurate 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. [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]

e. The law's choice.—My concern in this article is primarily to unearth what the law actually tells its factfinders to do. In its instructions, it 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. My thesis is that the law thereby treats all measures of truth simply as fuzzy sets. Whether the issue is occurrence or blameworthiness of an act, the law deals only in degrees of truth.[69]

I submit that the law has chosen to speak in terms of degrees of membership, because they behave more appropriately than classical probabilities in a world of imprecision.[70] Imagine the law is trying to determine if Tom was at fault. A number of features of this issue of fault indicate that the 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 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 as a 1 or a 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 imprecise for five reasons: "Our evidence is 'never complete', is usually 'inconclusive', is frequently 'ambiguous', is commonly

---

[68]Kosko, supra note 11, 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]Although the law seems to treat all measures of truth simply as type-1 fuzzy sets, naturally enough it does so without adopting all the formal aspects of fuzzy logic. The law built on the intuition that partial truths exist, but never adopted all the complications that logicians have refined since 1965. That is, the law generally leaves the assessment of evidence to its factfinders' instinctive treatment. Thus, there is here no reason to explore the sophistication of the aforementioned logical operators for ultrafuzzy sets.

Similarly, if a litigated case entailed two or more completely independent elements that were perfectly crisp, but subject to random uncertainty, a very limited role for probability's product rule arguably could emerge. But again the law apparently chooses to ignore this rare complication. It instead seems to treat all measures of truth as equivalent, namely, as imprecise degrees of truth. See infra Part II-B (discussing the conjunction paradox).

[70]See Shapira, supra note 24, at 1614-15 (arguing for the superiority of degrees of belief).

'dissonant' to some degree and comes to us from 'sources having imperfect credibility'."[71] Fourth, the factfinder 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.

Among other advantages of degrees of membership standing as beliefs, they better capture the effect of imperfect evidence, which was the rallying cry of Baconian theorists.[72] 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,[73] and to the frequent *cris de coeur* of theorists who lament any intrusion of probabilistic mathematics into the very human process of proof.[74]

Even if in the end it remains up in the air as to which system, fuzzy logic or probability, would be the superior vehicle for the legal proof process, my real interest here is in which system the law actually chose. This is a descriptive rather than prescriptive question: which system better expresses the law's scale of likelihood of truth, which better explains the legal treatment of the well-known conjunction paradox, 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.

I am not saying that, accepting the relevance of multivalence, we need to abandon the probability idiom. It can be quite expressive in the realm of standards of proof. But when plunging to the depths, we need always to remember that the legal foundation is ultimately fuzzy in nature.

---

[71]Schum, supra note 34, at 216 (quoting Rudolf Carnap, An Introduction to the Philosophy of Science 51 (Martin Gardner ed., 1966)).

[72]See L. Jonathan Cohen, The Probable and the Provable 49-57, 245-64 (1977) (developing, as an alternative to Pascalian (or mathematicist) probability, a Baconian (or inductive) theory of probability), reviewed by David A. Schum, 77 Mich. L. Rev. 446 (1979); Carl G. Wagner, 1979 Duke L.J. 1071. Baconian theory tries to look not only at the evidence presented, but at the evidence not available. It makes *evidential completeness* a key criterion. It thereby stresses an important concern, but its other contributions to understanding standards of proof are insufficient to justify its complexity. See Ronald J. Allen, A Reconceptualization of Civil Trials, 66 B.U. L. Rev. 401, 422-23 (1986).

[73]See Clermont & Sherwin, supra note 30; Kevin M. Clermont, Standards of Proof in Japan and the United States, 37 Cornell Int'l L.J. 263 (2004); Clermont, supra note 28; Richard W. Wright, Proving Facts: Belief Versus Probability, in European Tort Law 2008, at 79 (Helmut Koziol & Barbara C. Steininger eds., 2009); Richard W. Wright, Proving Causation: Probability Versus Belief, in Perspectives on Causation 195 (Richard Goldberg ed., 2011).

[74]E.g., Jaffee, supra note 41, at 934-51 (attacking the use of probability in analyzing proof); Tribe, supra note 3 (writing the classic version of the lament); 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").

Here is a last clever image to make the point as to what the law is interested 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.[75]

B. Legal Application: Gradated Likelihood

The way to apply fuzzy logic to factfinding 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 factfinder believes the element to be. One could express $\mu(x)$ as truth(x). Membership thereby creates degrees of truth.

While this membership will turn on likelihood of truth in a sense, it is in a sense different from the factfinder's subjective probability that the element is true. As already explained, subjective probability crisply deals with the "provably probable," while the fuzzy logic vaguely deals with the "probably provable."[76] 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.[77] For a significant example, the law today seems to limit 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 acknowledged that the conceivable spectrum

---

[75] Bogdan R. Kosanovic, Fuzziness and Probability, http://www.neuronet.pitt.edu/~bogdan/research/fuzzy/fvsp/fvsp.html (1995).

[76] See Schum, supra note 22, at 243 (crediting Judea Pearl, Probabilistic Reasoning in Intelligent Systems 571 (1988), for this phrasing).

[77] 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).

of standards had coalesced irresistibly into three.[78] For another example, the harmless-error doctrine frequently speaks of various low possibilities of an error's effect on outcome.[79] More generally, the law's standards of decision invoke a gradated scale of likelihood stretching across the broader spectrum from the slightest possibility up to virtual certainty.

The reason for this coarse gradation, I thought, lay in the cognitive psychology literature. Studies of humans' absolute judgment, short-term memory, and use of biased heuristics all supported the limited capability of humankind. Those studies suggested that a limited, step-like scale accords with how humans naturally process information. Judged likelihood usually finds expression in terms of a very small set of broad categories, which tend not to be quantified but rather to be expressed verbally. Moreover, cognitive limitations leave humans able only weakly to judge likelihood. It might therefore be more psychologically feasible to ask factfinders for a degree of truth than for their view of probability.[80] Perhaps the law, by an intuitive but wise reconciliation with cognitive limitations, had optimized by 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.[81]

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] So, I adhere to my prior view. But I might now state it in fuzzier terms. I would clarify my quoted contention so as not to imply that factfinders' gradated output expresses only random uncertainty, but instead to say that

---

[78]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) (similar).

[79]See Clermont, supra note 20, at 1121-23.

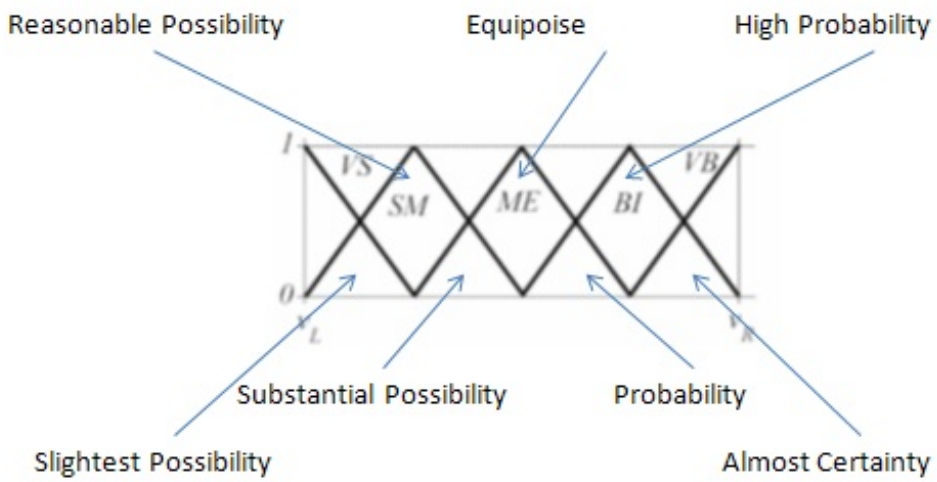
[80]See id. at 1139-40 (recounting empirical evidence); Yanlong Sun, Hongbin Wang, Jiajie Zhang & Jack W. Smith, Probabilistic Judgment on a Coarser Scale, 9 Cognitive Sys. Res. 161 (2008) (recounting more recent, consistent results).

[81]Clermont, supra note 20, at 1156.

[82]See Terrence L. Fine, Comment, 1 Stat. Sci. 352, 353 (1986).

their gradated output also expresses imprecision.

Thus, fuzzy logic accommodates the cognitive limitations of humans. Its real value, however, is that it captures the epistemic function in law better than probability. Fuzzy logic offers a rich approach to explaining and elaborating the law's gradated scale of likelihood. 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. For that reason, I espouse viewing the seven gradations as degrees of truth in the broad sense of fuzzy logic:



II. Combining 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

1. Maximum and Minimum

The power of this conceptualization becomes more obvious when one considers the combination rules of fuzzy logic. These can be quite complicated, but for our purposes those of special interest are the simple fuzzy operators or connectives for intersection and union of fuzzy sets,

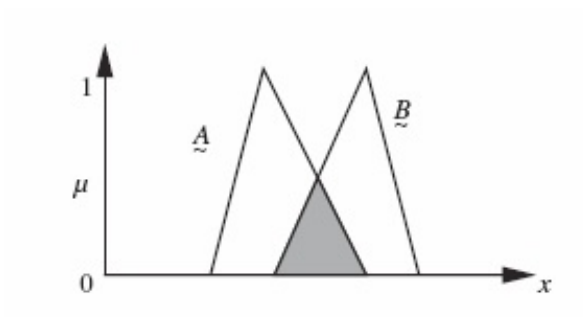
along with negation. Typically, these three operators are specified to define any logical system.[83]

Let X be the universe, and let $\underset{\sim}{A}$ be one fuzzy set and $\underset{\sim}{B}$ be another fuzzy set therein. 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. Designate the membership of element x in $\underset{\sim}{A}$ as truth(x), and the membership of element y in $\underset{\sim}{B}$ as truth(y). The AND and OR functions for intersection and union in fuzzy logic work this way:

$$\text{truth}(x \text{ AND } y) = \text{minimum}(\text{truth}(x), \text{truth}(y))$$

$$\text{truth}(x \text{ OR } y) = \text{maximum}(\text{truth}(x), \text{truth}(y))$$

The following diagram may help to visualize the former function, the 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. For any one x that falls in the intersection, the degree of membership therein will be the degree of membership in $\underset{\sim}{A}$ or $\underset{\sim}{B}$, whichever has the lower membership line at that point x:[84]



Or let X be the universe of men, and let $\underset{\sim}{A}$ be the set of tall men and $\underset{\sim}{B}$ be the assumedly independent set of smart men. So, if Tom is a .30 member of $\underset{\sim}{A}$ and a .40 member of $\underset{\sim}{B}$, 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 $\underset{\sim}{A}$ and $\underset{\sim}{B}$.

2. Product Rule Contrasted

a. Different rules.—The reader should sense that something odd is afoot. After all, the

---

[83]See Sider, supra note 7, at 67-70. For elaborations of fuzzy intersection and union, see Radim Bělohlávek, George J. Klir, Harold W. Lewis III & Eileen Way, 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).

[84]The figure comes from Ross & Parkinson, supra note 18, at 33.

probabilistic operation for AND is multiplication of the probabilities of independent events.[85] But fuzzy logic tells us to apply the MIN operator even for independent events.

Think of a room with ten men each 5'6" tall. We might think of each as .30 tall. What would we term the tallness of the ten men together? 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.

Now if the room has ten men, with three "short" men 5'6" or below and three "dumb" men, then one man picked at random has a .09 chance of being both short and dumb. So, here the product rule applies.

That is, the probabilistic operator is not inconsistent with the fuzzy operator. Instead, the two apply in different contexts.

Where the contexts overlap, the two operators are equivalent.[86] 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 true or 0 if not. That is, the product rule would be no different from the MIN operator: $\text{truth}(x)\cdot\text{truth}(y)=\text{minimum}(\text{truth}(x), \text{truth}(y))$.

The product rule takes on a special role only when the world is fuzzy, but the values of x and y happen to be randomly uncertain without being vague. It makes 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.[87] The chance of conjunction, that x and y will both turn out to be true=1, will be the product of the individual probabilities.

Because both the product rule and the MIN operator can give correct, but sometimes different, answers, they must be giving answers to different questions. The conjunction of probabilities is answering a different question than the intersection of degrees of truth is answering. The conjunction of probabilities gives the chance of something, 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, while the product rule is the special way to handle random uncertainty for independent

---

[85]Analogously, De Morgan's law 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. But fuzzy logic tells us to apply the MAX operator.

[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 postulate of the excluded middle).

[87]Yet, applying the product rule to subjective probabilities sometimes seems illogical. 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 51, at 375, 403.

events in the crisp world of classical probability.

b. Choosing the appropriate rule.—Which is the correct question to ask in a particular situation, whether it arises in a crisp or fuzzy world? The nature of the desired answer will determine whether the product rule or the MIN operator is appropriate.

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 $\underset{\sim}{A}$ and a .40 member of $\underset{\sim}{B}$. Those numbers would read in fuzzy terms 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 person." The MIN operator yields a belief in that intersection of .30.

The probabilistic calculation, however, would yield a statement like "Because Tom is not so tall and Tom is not so smart, Tom is likely a short, dumb person." The chance of a tall and smart Tom according to the product rule are .12, not .30. The probabilistic result 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 persons. Multiplication of probabilities gives the chance that Tom is both completely tall and completely smart, while what we may 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 people. The probabilistic calculation will approach .00, even if some of the values are very high. The fuzzy combination will go no lower that 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[88] and some pinkish hue, one would give it, perhaps, a .6 membership in the 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 chairness 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.

Third, picture a column of 100 coins, 30 of them heads randomly placed, and another column of 100 coins, 40 of them heads randomly placed. Then only about 12 paired rows will be two heads. Or picture a column of 100 people, 30 of them tall people randomly placed, selected from a universe where people are either completely tall or completely short; and picture another column of 100 more people, 40 of them smart people randomly placed, selected from a universe where people are either

---

[88]See H.G. Wells, A Modern Utopia 381 (1905) ("I would undertake to defeat any definition of chair or chairishness that you gave me").

completely smart or completely dumb. Then only about 12 paired rows will be tall and smart, respectively. Now, picture instead a column of varying beliefs in the tallness of 100 people selected from a universe where people have the tallness trait distributed naturally, aligned from tall down to short, and another column of 100 beliefs, 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.—Degrees of truth and classical probability are undeniably different. They behave differently in the conjunction or intersection setting. Once lawmakers have in mind the difference between the product rule and the MIN operator, they have to decide which the law should apply. Which do they want to use in a factfinding endeavor? Is the law interested in the "provably probable" or in fuzzy logic's "probably provable"?

Imagine the law is trying to determine if Tom 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, the law should be even less concerned with the chance of a 1 repetitively turning up; applying the product rule to subjective probabilities here seems illogical. In short, we want to know if it is a reddish chair, not what the chances are that it is 100% a chair and 100% red.

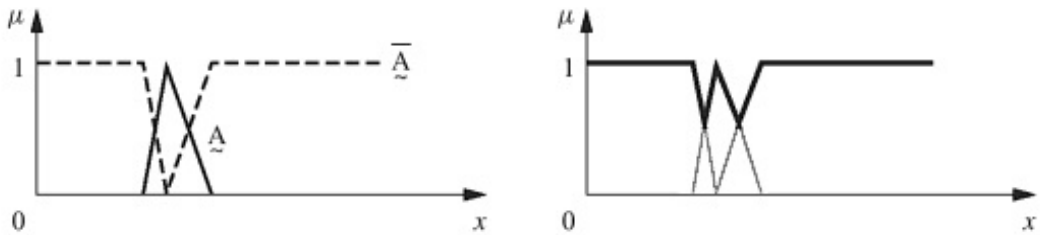
I therefore submit that degrees of truth behave more appropriately than classical probability in a world of imprecision. But, again, my central question is which representation the law employs. Here, as I have already said, and as I shall show in the upcoming resolution of the conjunction paradox, I am confident that it is degrees of truth.

3. Negation Operator

Accepting the usefulness of fuzzy logic prompts interest in other fuzzy operators. Another basic one is negation. Here truth(not$x$)=(1 − truth(x)), just as in standard logic. The negation is the complement.

However, the whole fuzzy set and its complement do not necessarily add to unity, reflecting that 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 $\underset{\sim}{A}$ by the solid line and its complement by the dotted line, and the right graph's dark upper line representing the union of $\underset{\sim}{A}$ and its complement by operation of the MAX function:[89]

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



The serrations in the upper line in the right-hand graph show that $\underset{\sim}{A}$ 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 changes the view of the proof process. Consider 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 both causation and fault and that these two elements are independent. 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 conjunction. 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 seriatim 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 element by assessing the conjunction of the evidence for it.[90]

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 result in liability.[91] 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."[92]

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. Or, because humans might tend to construct a story for the whole case rather than proceeding element-by-element, the factfinder might end up applying the standard of proof to the conjoined elements.

In fact, many psychologists agree that the factfinder naturally constructs such stories, although

---

[90]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).

[91]See James A. Henderson, Jr., Fred Bertram & Michael J. Tõke, Optimal Issue Separation in Modern Products Liability Litigation, 73 Tex. L. Rev. 1653, 1655-59, 1667-75 (1995).

[92]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.

perhaps not in a very systematic manner.[93] The broadly accepted *story model* of evidence processing holds that the factfinder, over the trial process's course, constructs from the evidence the story that makes maximal sense; and the factfinder then chooses, among the available decisions, the one that fits best with the story it has constructed:

> 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.[94]

If a juror constructs a story for the whole case, or otherwise cognitively processes 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 factfinder 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, this may work only to encourage the jurors' detailed evaluation 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

---

[93]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, Steven Penrod & Nancy Pennington, Inside the Jury 22-23 (1983) (providing a brief summary of empirical studies supporting the story model of evidence processing); Paula L. Hannaford, Valerie P. Hans, Nicole L. Mott & G. Thomas Munsterman, 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: The Psychology of Juror Decision Making 192 (Reid Hastie ed., 1993) (detailing the story model and summarizing empirical studies testing it).

[94]Frederick, supra note 93, at 296-97 (citations omitted).

brake any undesirable tendency of the factfinder to rush toward creating a story.[95]

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.[96] The conjunction paradox therefore remains troubling—but only if the law really calls for violating an applicable product rule.

The law, however, does not. Instead, the conjunction paradox implodes under the force of fuzzy logic. 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. 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 nonliability combined.

Comfortingly, applying the standard of proof element-by-element works out to be equivalent to applying it to the whole conjoined story, and the proof process within elements is equivalent to the proof process between elements. So, if the factfinder really does follow the story model, that practice would not directly endanger the standard of proof.

The law does seem to know what it is doing, then. By phrasing its instruction to apply the standard of proof element-by-element, it is instructing to apply the MIN operator. But do actual factfinders 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.[97] Nevertheless, no sign exists that factfinders in the legal system are

---

[95]See Bruce Ching, Narrative Implications of Evidentiary Rules, 29 Quinnipiac L. Rev. 971 (2011).

[96]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 decisionmaking practice).

[97]See Gregg C. Oden, Integration of Fuzzy Logical Information, 3 J. Experimental Psychol.: Hum. Perception & Performance 565 (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. In any event, in the legal system any human failing to apply the MIN operator would be offset by the human tendency to construct a story for the whole case instead of

using the product rule, an absence that generated the unfounded fear of the conjunction paradox in the first place.[98]

III. Analyzing Beliefs

This Part will introduce belief functions into the mix, in order better to represent how imperfect evidence keeps factfinders 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 factfinders at point zero.

A. Shafer's Belief Functions

1. Basics of Theory

I have already implicitly suggested treating the degree of S's truth, which is a degree of membership in the set of true facts, as instead a degree of belief in S as a true proposition. The broad version of the theory of belief functions will then give us a handle on how to treat such beliefs.[99] It will also provide us with a better mental image for representing indeterminacy.[100]

---

proceeding element-by-element.

[98]Another claimed paradox involves multiple theories, in which the law denies relief to a supposedly deserving plaintiff:

> 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.

Ronald J. Allen & Sarah A. Jehl, Burdens of Persuasion in Civil Cases: Algorithms v. Explanations, 2003 Mich. St. L. Rev. 893, 939 (citing Saul Levmore, Conjunction and Aggregation, 99 Mich. L. Rev. 723, 745-56 (2001)). 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. See supra note 85.

[99]See Shafer, supra note 51, at 35-37.

[100]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,

I draw on the belief function theory's description of a single factual hypothesis, which creates a so-called power set of four beliefs: S, notS, neither S nor notS, and either S or notS. In factfinding, I 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 $0 \leq \text{Bel}(S) \leq 1$.

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. But you think that one saw a different woman's body, which discounts the evidence of death but gives no support to her being alive. And you think that another was lying as part of a coverup of her escape from captivity, which is compatible with both S and notS 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 tied to 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. The belief in the "null" of neither alive nor dead is set by definition to be 0. The belief in the catchall "either," alive or dead, is 1.0.

```
|            Bel(S)            |                | Bel(notS) |
|++++++++++++++++++++++++++++++|________________|-----------|
 <----   Plausibility of Katie's death   ---->
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.[101] 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. (The traditionally expressed probability of her being dead would fall within the range from the lower to the upper probability.)

The plausibility that she is alive totals .5, being .2 + .3. Plausibility numbers each equal one minus the belief in the opposite.

Belief functions thus harness the idea of imprecise probability to capture indeterminacy. They make an independent contribution to understanding law. They can be used with expressions of

---

supra note 51, at 1, 2-19 (recounting also the history of belief function theory).

[101] See Barnett, supra note 51, at 200-01 (providing a neat mental image for the 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).

probability. But combining belief functions with fuzzy logic's degrees of truth and its operators makes an even bigger step toward that understanding.[102]

These beliefs can be expressed, if expression is ever necessary, as coarsely gradated beliefs. In addition to 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 could 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.

2. Negation Operator

By classical 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, as normally some belief remains uncommitted. Thus, for Katie, Bel(S)=.5 and Bel(notS)=.2, so the sum of determinate beliefs adds to .7.

3. Lack of Proof

Classical probability encounters legendary difficulties with a state of ignorance.[103] The reason is that it cannot distinguish between lack of belief and disbelief. In Pascalian 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.[104] 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 probability, and it produces inconsistencies when there are more than two hypotheses in play.

Meanwhile, one of the great strengths of belief function theory is that it well represents a state

---

[102]See Dubois & Prade, supra note 87, 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 51, 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).

[103]See, e.g., Richard Lempert, The New Evidence Scholarship: Analyzing the Process of Proof, 66 B.U. L. Rev. 439, 462-67 (1986).

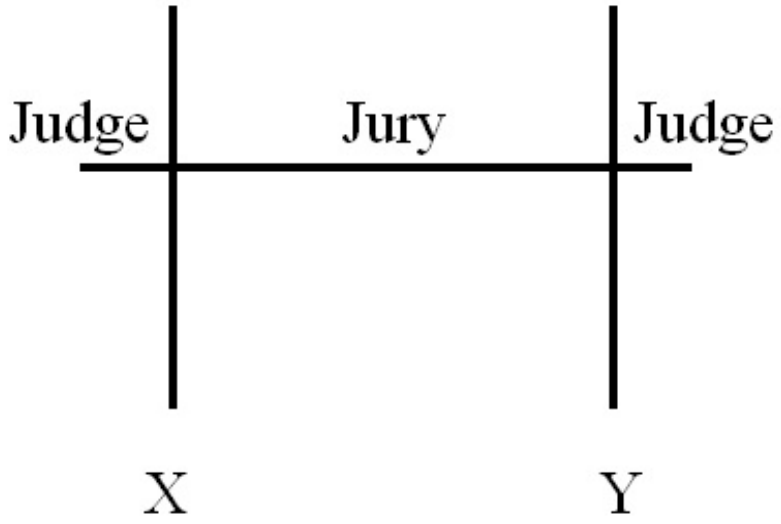
[104]See Lea Brilmayer, Second-Order Evidence and Bayesian Logic, 66 B.U. L. Rev. 673, 686-88 (1986).

of ignorance.[105] An inquirer, if ignorant, starts at zero, not at a 50% belief. When Bel(S)=0, it does not mean 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, the fact is that Bel(notS)=0 too. The uncommitted belief is the entirety or 1, meaning that S is "plausible." In other words, we do not believe or disbelieve S.

Thus, belief function theory makes a very useful distinction between disbelief and a lack of belief, or disproof and a lack of proof. After all, disbelief and a lack of belief are entirely different states of mind.

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 factfinder 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 factfinder of the truth of an element or suffer adverse determination on that element.



[105]See Shafer, supra note 51, at 22-24 ("representation of ignorance").

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.[106] 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 factfinder, not the judge's thought process if the judge were finding facts. Alternatively stated, this diagram represents the burden of production, not the burden of persuasion.

The plaintiff in the imagined case starts at the left of the diagram, at the zero point. 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 beyond line X 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[107]) describes an inference that a jury 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

---

[106]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).

[107]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 doctrine by use of fuzzy logic).

jury zone between Y and X.[108]

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,[109] on mere disbelief of the opponent's testimony,[110] or on demeanor evidence drawn from the opponent's testimony.[111] Similarly, naked statistical evidence normally will not satisfy the initial burden of production.[112] 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.[113] 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.[114] In the absence of these special rules, any burdened 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 scheme seems fairly compatible with classical probability. One qualification is that representing the judge's view of jury error as a fuzzy degree rather than a point probability would better capture reality.

But the biggest difficulty for probability is explaining why the law starts at zero rather than

---

[108]See Fed. R. Evid. 301.

[109]See Stimpson v. Hunter, 125 N.E. 155 (Mass. 1919).

[110]See Cruzan v. N.Y. Cent. & Hudson River R.R. Co., 116 N.E. 879 (Mass. 1917).

[111]See Dyer v. MacDougall, 201 F.2d 265 (2d Cir. 1952).

[112]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 nonparadoxical 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 factfinder converts statistical evidence into a belief).

[113]See, e.g., Baxter v. Palmigiano, 425 U.S. 308, 316-20 (1976) (treating failure to testify).

[114]See Robert S. Summers, Evaluating and Improving Legal Processes—A Plea for "Process Values," 60 Cornell L. Rev. 1 (1974).

at 50%. The probabilist might argue 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%.[115] Then, if the plaintiff offers a feather's weight of evidence, he should 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 a feather's weight of evidence, and does so by summary means. Why is that? The law says that we should start at zero, and to get to X requires more than a feather's weight.

In starting at zero, the law seems to be proceeding under belief function theory. Thereby, 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.

IV. Applying Standards

This Part will introduce the idea of comparing belief and disbelief of a fact, when done 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 traditional academic view of the proof process rests on a two-valued logical approach that the law does not and should not employ. Factfinders 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 factfinders 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 factfinders back into what looks like a two-valued logic, by forcing them to decide for one party or the other.[116] Such disambiguation is not a

---

[115] See Anne W. Martin & David A. Schum, Quantifying Burdens of Proof, 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%).

[116] A separable question is whether the law should instead deliver partial relief following partial proof. 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 economic superiority of current law to an expected-value approach that would award damages proportional to probabilistic certainty). After cataloging the prevalence of fuzzy concepts in law, Katz, supra note 63, at 157-81, concludes that law is correct to draw lines

practice unique to law. All fuzzy computer programs end with a step that produces an unambiguous output, a step called *defuzzification*.[117]

Application of a standard of proof is a different step from evidential inference or argument; the academic disputes as to standards do not overlap with the disputes over how to assess and analyze evidence.[118] Psychologists have contributed almost nothing here,[119] 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 here too. Admittedly, one could pursue the image of classical probability, a two-valued logic. Then the ultimate task of applying a standard of proof would unavoidably involve placement on a scale of likelihood.

A better understanding of standards of proof would come from thinking in terms of many-valued logic and belief functions, however. The law does not need to require evidence making the fact more likely than 50%. Even though decisionmaking requires converting from a many-valued logic to an output that sounds two-valued, the decision might involve only comparing Bel(S) and Bel(notS) while ignoring the indeterminate belief.[120] By requiring only a comparison, belief functions would never require placement on a scale of likelihood. All the factfinder need do is compare the strengths of belief and disbelief.

B. Legal Application: Burden of Persuasion

---

in the fuzz and so separate all-or-nothing remedies. Professor Katz argues that law must establish discontinuities in order to retain its rational coherence.

[117]See Kosko, supra note 11, at 172 (describing fuzzy computer systems).

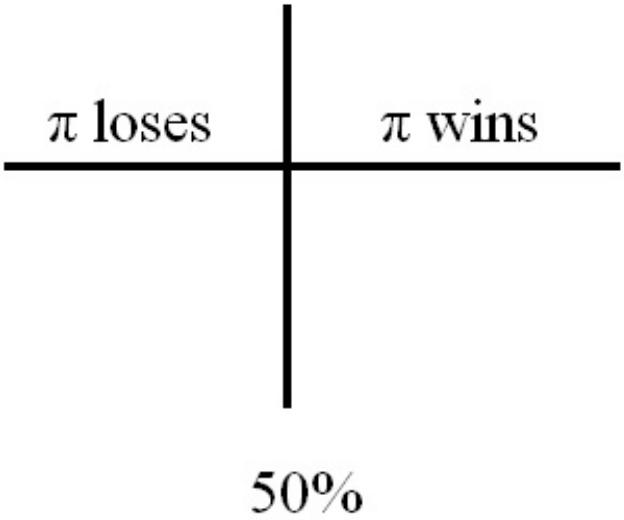
[118]See Peter Tillers & Jonathan Gottfried, Case Comment—United States v. Copeland, 369 F. Supp. 2d 275 (E.D.N.Y. 2005): A Collateral Attack on the Legal Maxim That Proof Beyond a Reasonable Doubt Is Unquantifiable?, 5 Law Probability & Risk 135, 142 (2007) (observing the difference between processing evidence and applying standards 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 decisionmaking 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 51, at 693.

[119]See Clermont, supra note 28, at 477, 485 (recounting the state of psychological knowledge on standards of proof).

[120]See Smith, supra note 65, at 503 (describing a betting scheme based on such comparison). For the mathematics involved, see Smets & Kennes, supra note 118, at 703-11.

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 factfinder in ultimately weighing the evidence. The grid now measures the factfinder'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, 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 factfinder thinks that he has not passed the 50%-line, then the factfinder 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 factfinder as evenly balanced.[121]

---

[121] 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 factfinder."); Clermont, supra note 20, at 1119 n.13, 1122 n.36, 1147-48; Neil B. Cohen, Conceptualizing Proof and Calculating Probabilities: A Response to Professor Kaye, 73 Cornell L. Rev. 78, 90-91 (1987); Neil B. Cohen, Confidence in Probability: Burdens of Persuasion in a World of Imperfect Knowledge, 60 N.Y.U. L.

This zone of equipoise means that the burden of persuasion will affect many more cases than those in which the conflicting evidence results precisely in a dead heat. Thus, the factfinder 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 risk of nonpersuasion of a fact—matters.[122] The plaintiff may appear to start from zero, or the defendant may. An anchoring heuristic lowers the willingness of the factfinder to determine that the burdened party has prevailed, because people fail to adjust fully from a given starting point, even if arbitrarily set.[123] In sum, the burden of persuasion is not a mere tiebreaker, explaining why lawyers and judges fight and suffer over it in practice.

Again, this diagrammatic scheme for the burden of persuasion seems fairly compatible with classical 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 outcome will be a conclusion that this diagram for the burden of persuasion is fundamentally misleading. The law does not and should not conform to the prevailing 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.[124] (3) The standard of *proof beyond a reasonable doubt* means proof to a virtual-certainty. It very rarely prevails outside criminal law.[125]

b. Relative-plausibility theory.—The insightful relative-plausibility theory of Professor Ron

---

Rev. 385, 418-19 (1985).

[122]See Eyal Zamir & Ilana Ritov, Loss Aversion, Omission Bias, and the Burden of Proof in Civil Litigation, 41 J. Legal Stud. (forthcoming 2012), available at http://ssrn.com/abstract=1868465, at 23-24 (Nov. 9, 2011).

[123]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 8x7x6x5x4x3x2x1, while another group faced 1x2x3x4x5x6x7x8; the first group's median estimate was 2250, while the other's was 512; the correct answer is 40,320. Id.

[124]See Santosky v. Kramer, 455 U.S. 745 (1982).

[125]See generally Clermont, supra note 20, at 1118-21.

Allen shows a nontraditional preference for comparison rather than absolute judgment of likelihood.[126] He builds on the story model of evidence-processing to produce another theoretical brand of holism. The relative-plausibility theory posits that the factfinder constructs the story (or stories) that the plaintiff is spinning and another story (or stories) that the defendant is spinning. The factfinder 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.[127] 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

---

[126]See Allen & Jehl, supra note 98, at 929-43 (summarizing Professor Allen's previous work on the theory).

[127]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) (admitting expert evidence based on the weight of the evidence approach: "The scientist 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*. Professor Allen is drifting in his thinking in this direction. See Ronald J. Allen & Michael Pardo, Juridical Proof and the Best Explanation, 27 Law & Phil. 223 (2008). But Larry Laudan, Strange Bedfellows: Inference to the Best Explanation and the Criminal Standard of Proof, http://ssrn.com/abstract=1153062, powerfully demonstrates that inference to the best explanation holds little additional promise of explaining or illuminating standards of proof.

preponderance of the evidence.[128] Its more obvious, and admitted,[129] difficulty is that it does not track well what the law tells its factfinders about how to proceed, and it diverges from the law by requiring the nonburdened party to choose and formulate a competing version of the truth. Finally, it comes with baggage, such as seeming to compel acceptance of the story model.

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 Bel(S) and Bel(notS), in any of a variety of ways.

Although one could conceivably say that Bel(S) must exceed 50%, this formulation does not accord with the import of real cases in the common-law world.[130] 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.

Accordingly, the better approach is not to look at just one of the two beliefs compared to some absolute measure, but instead to compare the two. In comparing them, Bel(notS) is the belief in the negation of S, not the complement of Bel(S). It represents how much the factfinder actively disbelieves S, the fact in dispute. As to that fact, a belief in the falsity of the burdened party's version of the truth may develop naturally, even upon hearing only the burdened party's evidence. Also, the nonburdened party's evidence should contribute to raising Bel(notS).

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).

---

[128]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). 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 factfinder "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) (saying that the prosecution must "show that there is no plausible account of innocence").

[129]See Ronald J. Allen, Standards of Proof and the Limits of Legal Analysis, http://ssrn.com/abstract=1830344, at 14 (May 3, 2011).

[130]Perhaps, with their emphasis on "conviction" in the *intime conviction* standard, the civil-law countries test belief on an absolute scale in this way rather than by comparing beliefs. The absolute judgment approach fit better with an inquisitorial model than it did with an adversarial model, allowing it to persist for centuries. But its survival to today may rest instead on the system's desire to enhance the appearance of legitimacy. See supra note 73.

This comparative approach to the civil standard of proof does not mean that the nonburdened party needs to formulate a competing version of the truth, other than negation. Some scholars worry that looking only at negation puts the burdened party in the impossible situation of disproving every alternative possibility.[131] But that worry comes from confusing lack of belief with disbelief. Disbelieving S involves the degree to which the factfinder 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. The comparison involves the belief in notS, and does not involve the plausibility of notS. 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.

Now, as to the other two standards of proof, clear and convincing evidence should mean Bel(S) >> Bel(notS).[132] This standard should 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.[133] Appellate courts use it in reviewing judge-found facts.[134] 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.

As to proof beyond a reasonable doubt, it should mean not only that Bel(S) >> Bel(notS) but

---

[131]See, e.g., Michael S. Pardo, Second-Order Proof Rules, 61 Fla. L. Rev. 1083, 1093-94 (2009) (footnotes omitted) (speaking of the comparison imposed by more likely than not, but using "negation" in the sense of the complement of Bel(S)):

> [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 with regard to the goals of error minimization and allocation. If the plaintiff must prove that some fact, is more probable than its negation, *not-X*, then the plaintiff should have to show not only *X*, 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.

[132]Cf. Laudan, supra note 127, at 8-9 (discussing attempts to append such a notion to the approach of inference to the best explanation).

[133]See Clermont, supra note 20, at 1126-28, 1152-56 (describing new trial practice).

[134]See id. at 1128-30 (describing appellate review).

also that no reasonable doubt persists. A reasonable doubt means that a reasonable person could hold Bel(notS) > 0.[135] On the view that anything is possible, zero as a coarsely gradated degree of belief equates to a "slightest possibility."[136] Therefore, Bel(notS) > 0 means that for a conviction the prosecutor must show that no "reasonable possibility" of innocence exists.

d. Compatibility of reformulated and current standards.—A reader always suffers the temptation, upon seeing what looks like a plea for reconceptualization, to dismiss it as a pie-in-the-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. 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 to little more than explain what 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 world. One way to do that would be 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 the my formulations of the standards 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

---

[135]See Allen & Leiter, supra note 128, at 1528 (saying that the prosecution must "show that there is no plausible account of innocence"); cf. Laudan, supra note 127, at 9-12 (discussing attempts to append such a notion to the approach of inference to the best explanation).

[136]"Most legal decision making, like that in many other areas of complex activity, is done under conditions of uncertainty." Michael J. Saks & Robert F. Kidd, Human Information Processing and Adjudication: Trial by Heuristics, 15 Law & Soc'y Rev. 123, 126 (1981). 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., e.g., Neal Gabler, The Elusive Big Idea, N.Y. Times, Aug. 14, 2011 (Sunday Review), at 1 (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., 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, 136, at 104 (rebutting "the myth of certainty").

conforms better to the actual law than the probabilistic view does.

Under the standards' reformulation, the evidence will support S to an extent while supporting notS to another extent, and the factfinder need only compare the two fuzzy beliefs. How does the current law actually formulate, say, preponderance of the evidence? Consider a couple of classic cases.

In *Livanovitch v. Livanovitch*,[137] 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."[138] The appellate court said:

> 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.[139]

In *Lampe v. Franklin American Trust Co.*,[140] 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."[141] 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."[142]

---

[137]131 A. 799 (Vt. 1926).

[138]Id. at 800.

[139]Id.

[140]96 S.W.2d 710 (Mo. 1936).

[141]Id. at 723.

[142]Id. (quoting Rouchene v. Gamble Constr. Co., 89 S.W.2d 58, 63 (Mo. 1935)).

These two courts' formulations seem contradictory. But if the quotations are interpreted as speaking in nonprobabilistic 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 say 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) "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.[143] My approach would conform to the middle option of (2), rather than either (1) relative plausibility or (3) absolute judgment.

In the end, I submit that comparison of coarsely gradated beliefs is an accurate representation of what the law tells a factfinder to do with a standard. In civil cases, the factfinder has to find that Bel(S) is more likely than *not*, which actually means Bel(S) > Bel(notS). Or as the judge tells the jurors, a preponderance means that the evidence "produces in your minds belief that what is sought to be proved is more likely true than not true."[144] By literally instructing factfinders 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 factfinders to form a set of fuzzy factual beliefs that conform to a coarsely gradated 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 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 factfinders at the end of a case would properly apply the standard to each separate element. It also implies that the factfinders should start the case, being in a state of ignorance and 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 quantified expressions, nicely explain the law's development of a coarsely gradated scale of possibilities and probabilities: (1) slightest possibility, (2) reasonable possibility, (3) substantial

---

[143]See, e.g., Covington, supra note 90, at 99-100.

[144]3 Kevin F. O'Malley, Jay E. Grenig & William C. Lee, Federal Jury Practice and Instructions: Civil § 104.01 (6th ed. 2011).

possibility, (4) equipoise, (5) probability, (6) high probability, and (7) almost certainty. And the coarseness of the scale of likelihood means that the factfinder in comparing beliefs will not have to draw paper-thin distinctions.

Second, when the factfinders face multiple elements, it has long appeared that they seek out 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, I can demonstrate the meshing of the burden of production with, say, the new view of the preponderance standard by imagining a directed verdict motion. It would require the judge to ask if a reasonable jury could view Bel(S) > Bel(notS). At the end of the plaintiff's case, if a reasonable Bel(notS) equals 0 (effectively a "slightest possibility"), then the inequality requires a compatibly reasonable Bel(S) to exceed 0 (effectively a "reasonable possibility"). The demand for showing a reasonable possibility is the embodiment of the burden of production, and it is what keeps the plaintiff from succeeding with a mere feather's weight of evidence. An exemplary situation would be where the plaintiff has produced a little evidence, but it is "pure" evidence that gives the defendant no support.[145] 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 in the rest of the trial, but moves again for a directed verdict at the end of all the evidence, then the case should go to the jury. The jury, if it takes the same view of the evidence as the judge hypothesized, would find for the plaintiff—even on such thin evidence.[146]

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 therefore call for comparison of 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 factfinders 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

---

[145]See Liu & Yager, supra note 100, at 18-19 (discussing Liebniz's notions of pure and mixed evidence).

[146]See Laudan, supra note 127, at 13-14 ("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 sufficiently good story to tell.' Under current rules, if the plaintiff has a better story than the defendant, he must win the suit, even if his theory of the case fails to satisfy the strictures required to qualify his theory as the best explanation.").

50% appeared appropriate for civil cases: among competing fixed standards,[147] it minimizes the expected number of erroneous decisions and also the expected sum of wrongful amounts of damages,[148] which is the goal that the law apparently pursues in preference to optimizing incentives

---

[147]If we knew more about the realities of the type of case or 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 value 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.

[148]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 worthy of consideration—the preponderance standard should perform better than any other nonvariable 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 minimizes the system's expected error costs. Let p be the apparent probability that the defendant is liable (for D dollars). If $p > ½$, call it p_1; and if $p \leq ½$, 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 $\sum[(1-p_1)D + p_2D]$. On the other hand, under a very high standard of proof that eliminates false positives, the analogous sum is $\sum[p_1D + p_2D]$. 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) (arguing that the $p>½$ rule is appealing because "it minimizes expected losses"); 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

for primary conduct.[149] 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. 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 $p > ½$. Call p by the name p_1 in that case; and if Bel(S) ≤ Bel(notS), 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 $\sum[(1-p_1)D + p_2D]$. On the other hand, under a very high standard that eliminates false positives, the analogous sum is $\sum[p_1D + p_2D]$. Therefore, given that $(1-p_1)$ is less than p_1, the reformulated preponderance standard lowers the system's expected error costs.

To close, perhaps a comprehensive example would be of benefit. Suppose that someone has seriously injured Katie, 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 factfinder might assess that evidence and conclude as follows: (1) The evidence points to Tom being the perpetrator. If the factfinder 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 is 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 factfinder 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 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 factfinder is pretty convinced nevertheless. He would put causation at .8, or highly probable. (4) Katie'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 factfinder's task as traditionally envisaged.
- Second, the factfinder may want to combine these findings. They are a mixture of probabilities and degrees of truth. But viewing them all as degrees of truth allows the use of the MIN operator, so that he can say that Katie's story comes in at .6, or probable. Katie should win, by the use of fuzzy logic.

U. Pa. L. Rev. 1159 (1983) (considering bias in the distribution of errors).

[149] 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 applications of the law. 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).

- Third, this approach does not do a terribly good job accounting for the state of the evidence. It still poses an odd question to the factfinder: given imperfect evidence, what is the degree to which the plaintiff is right? Belief functions work better here to reflect the factfinder'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 that invoking the simplistic scale of likelihood just described. Instead of saying that Tom's fault is probable, the factfinder should speak and think in terms of degrees of belief. Here, although belief functions do not require placement on a scale, he in effect might end in believing Katie's position to be only substantially possible.
- Fourth, that situation does not mean that Katie will lose, however. All the factfinder must do is compare belief and disbelief. All that preponderance of the evidence requires is that the strength of the factfinder's belief that Tom was at fault must exceed his belief that Tom was not at fault. While some of the factfinder's belief remains uncommitted, he did find Katie's position to be a good one: more likely true than false. He might, if forced to express likelihood, believe to a merely reasonable possibility in the falsity of Tom's fault. So, Katie should win on the issue. Belief functions thus add the idea that the factfinder in such a case must have a belief in the case's truth stronger than his belief in its falsity.

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 traditional probabilistic view, this conceptualization improves our mental image of the factfinders' task, conforms more closely to what we know of people's cognition, and indeed captures better what the law says its standards are and how it manipulates them. But the conceptualization leaves the law's standards essentially intact to accomplish their current purposes.

In brief, the new logic reveals that the law wants factfinders to form degrees of belief that conform to a coarsely gradated 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. One virtue of the conceptualization is that it is not a radically different conception, as it principally acts to confirm the law's ancient message that factfinders should simply compare their nonquantified views of the fact's truth and falsity. 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 subjective probabilities.