Northwestern University Law Review
Printed in U.S.A.
Vol. #105 No. 1

THE FLAWED PROBABILISTIC FOUNDATION OF LAW AND ECONOMICS

Alex Stein[*]

INTRODUCTION

Law does not just tell people what is and is not allowed; it also informs them about penalties and rewards that attach to particular prohibitions and prescriptions.[1] These penalties and rewards guide individual decisionmaking. When a person weighs the costs of taking an action the law favors, the reward she can expect may make it worthwhile for her to take the action by raising her aggregate benefit above the sum of the associated costs. Conversely, when a person contemplates an action the law disfavors, the accompanying legal penalty imposes a cost that may wipe out the action's net benefit to the person and, with it, the action's appeal. The goal of these pe-

[*] Professor of Law, Benjamin N. Cardozo School of Law. I thank Ron Allen, Lior Avnaim, Rick Bierschbach, Shirley Blaier-Stein, David Carlson, Albert Choi, Bob Cooter, Shachar Eldar, Yuval Feldman, Dov Fox, Jim Gibson, Gabriel Halevy, Amal Jabareen, Sagit Leviner, Daphna Lewinsohn-Zamir, Shahar Lifshitz, Alan Miller, Gideon Parchomovsky, Ariel Porat, Fred Schauer, Joshua Teitelbaum, Avishalom Tor, and participants in workshops and presentations at the American Law and Economics Association Twentieth Annual Meeting, Bar-Ilan Law School and Ono Academic College, Israel, for insightful discussions and comments, and Stephen McKeown (J.D., Yale Law School, 2010) for outstanding research assistance during my visit at Yale in 2008–2009.

[1] *See* Robert Cooter, *Prices and Sanctions*, 84 COLUM. L. REV. 1523, 1524–31 (1984).

nalties and rewards is to align people's selfish interests with the interests of society. The law tries to make it privately beneficial for individuals to behave in socially desirable ways.

To affect individuals' choices among different courses of action, the law's threats of penalties and promises of rewards must be effectively implemented. The effectiveness of a threatened penalty depends on a person's probability of actually being penalized for taking an action disfavored by the law. By the same token, the effectiveness of a promised reward depends on a person's probability of actually being rewarded for taking an action that the law favors. This dependency is crucial. Because of informa-informational asymmetries and the high costs of law enforcement, the legal system often fails to deliver penalties and rewards to individuals who ought to receive them.[2] The law therefore does not really tell a person,"If you act in such and such a way, you will receive such and such penalty (or reward)." Rather, it tells a person,"If you act in such and such a way, you *probably* will receive such and such penalty (or reward)."

Hence, it is crucial to determine what "probably" means. This inquiry is fundamental to understanding the operation of legal rules and institutions. These rules and institutions form a system that incentivizes individuals to account for their probability of receiving the appropriate penalty (or reward) as a consequence of doing something that the law proscribes (or favors). What criteria do individuals use for determining probabilities that matter to them? What criteria should they use? Are these criteria similar to those upon which the legal system models its incentives for individuals' actions? These questions define the probability issue that this Article attempts to resolve.

The academic literature that examines the effects of legal incentives on individuals' actions is rich, heterogeneous, and insightful. Yet, it has never addressed the probability issue. Instead, it assumes that this issue is settled. According to this literature, a rational person has only one way of determining the probability relevant to her decision. The person uses her and other individuals' experiences to calculate or intuit the number of cases in which the legal system penalizes (or rewards) people in situations similar to hers and the number of cases in which it fails to do so. She then divides the number of cases in which the penalties (or rewards) are delivered by the total number of observed or intuited cases. The result of this calculation gives the person the probability in which she is interested. If she finds, for instance, that the legal system penalizes only half of the people who engage in a certain illegal action—say, running a red light—her probability of being penalized for taking a similar action will equal 0.5. The person will

[2] *See, e.g.*, Ehud Guttel & Alon Harel, *Uncertainty Revisited: Legal Prediction and Legal Postdiction*, 107 MICH. L. REV. 467, 468 (2008) (attesting that law enforcement is riddled with uncertainties).

then discount the law's penalty by 50%[3] and reach her expected cost of disobeying the law. If the full penalty for the wrongdoing is, say, a $1000 fine, the discounting willbring the expected penalty amount down to $500. The reduced amount might make it privately beneficial for the person to break the law. To fix this misalignment between the person's selfish interest and society's benefit, the legal system has to either enhance its enforcement efforts or double the fine.[4]

This traditional account of legal incentives postulates that individuals base their probability calculations upon "instantial multiplicity." The instantial multiplicity criterion associates probability with an event's frequency[5] or propensity,[6] whether observed[7] or intuited.[8] It encompasses two basic propositions. First, an event's chances of occurring are favorable when it falls into the majority of the observed or intuited events. Second, an event's chances of occurring are not favorable when it falls into the minority of the observed or intuited events.

These propositions are not tautological. They do not merely restate the numbers of relevant events that the reasoner counted or intuited. These propositions about an event's chances of occurring make a substantive epistemic claim about the reasoner's situation. They hold that the reasoner's consideration of past relevant events warrants an inference about what will happen in the action that she is presently considering. These propositions use instantial multiplicity to produce knowledge that did not exist before. According to this knowledge, the outcome of the reasoner's actions is most likely to feature the characteristics that belong to the mathematical majority of the previously observed or intuited events.[9]

---

3 If the person is averse towards risk, she will discount the penalty by a lesser amount. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW 10–11 (7th ed. 2007) (explaining neutrality and aversion towards risk).

4 *See* Gary S. Becker, *Crime and Punishment: An Economic Approach*, 76 J. POL. ECON. 169 (1968) (the foundational account); *see also* Richard Craswell, *Damage Multipliers in Market Relationships*, 25 J. LEGAL STUD. 463, 466 (1996) (noting importance of the idea that increasing fines while reducing enforcement maintains the overall deterrent effect); A. Mitchell Polinsky & Steven Shavell, *Punitive Damages*, *in* 3 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 192, 193–94 (Peter Newman ed., 1998); A. Mitchell Polinsky & Steven Shavell, *Punitive Damages: An Economic Analysis*, 111 HARV. L. REV. 869, 896–97 (1998) [hereinafter Polinsky & Shavell, *Economic Analysis*] (arguing that courts should take defendants' probability of escaping liability into account when calculating punitive damages).

5 *See* L. JONATHAN COHEN, AN INTRODUCTION TO THE PHILOSOPHY OF INDUCTION AND PROBABILITY 47–53 (1989) (discussing frequency-based probability).

6 *Id.* at 53–58 (discussing propensity-based probability).

7 *Id.* at 47–48 (stating the observational basis of frequency-based probabilities).

8 *Id*. at 58–70 (discussing probabilistic formulations of individuals' degrees of belief); *see also* D.H. MELLOR, THE MATTER OF CHANCE 1–18 (1971) (analyzing intuitively-formed personalist probabilities as beliefs in propensities and frequencies of events); Itzhak Gilboa et al., *Probability and Uncertainty in Economic Modeling*, 22 J. ECON. PERSP. 173, 175–82 (2008) (critiquing economic models that rely on subjective probabilities).

9 *See* COHEN, *supra* note 5, at 13–27 (outlining methods of inference based upon instantial multiplicity).

The instantial multiplicity criterion forms the basis of the mathematical probability system. All studies of law and economics accept this system as correct.[10] Court decisions dealing with the formation of legal incentives echo this academic consensus.[11] Neither lawyers nor economists have questioned the validity or applicability of the instantial multiplicity criterion. They simply accept this criterion and the resulting system of probability as intuitively appealing and operationally feasible.[12] In what follows, I call this system the "axiomatized view of probability" or, in short, the "axiomatized view."

The only challenge to the axiomatized view has been raised by psychologists and behavioral economists. These scholars accept the axiomatized view of probability as normatively correct but dispute its applicability to real-world decisions that ordinary people make about their affairs. Specifically, they claim that ordinary people often ignore base rates,[13] undervalue the probabilities of nonexperienced events,[14] overestimate the

---

[10] *See, e.g.*, GUIDO CALABRESI, THE COSTS OF ACCIDENTS: A LEGAL AND ECONOMIC ANALYSIS 250–51, 255–59 (1970) (criticizing the conventional, case-by-case method of ascribing liability for accidental damages and advocating transition to statistical models); POSNER, *supra* note 3 (relying on mathematical probability in all discussions throughout the book); STEVEN SHAVELL, FOUNDATIONS OF ECONOMIC ANALYSIS OF LAW (2004) (same); John E. Calfee & Richard Craswell, *Some Effects of Uncertainty on Compliance with Legal Standards*, 70 VA. L. REV. 965, 969–70 (1984) ("[D]efendants do not face a simple choice between actions certain to lead to liability and actions bearing no risk of liability at all. Instead, each possible action is accompanied by an associated probability that a defendant will be tried, found liable, and made to pay damages or a fine. . . . To the extent that defendants are influenced by the fear of liability, their behavior will be influenced by this distribution of probabilities, rather than simply by the nominal legal standard. Indeed, from the defendant's point of view the rule of law is that distribution of probabilities." (footnote omitted)); Gillian K. Hadfield, *Weighing the Value of Vagueness: An Economic Perspective on Precision in the Law*, 82 CALIF. L. REV. 541, 542 (1994) ("A law, from an individual's point of view, represents the compilation of that individual's assessment of the probabilities of being held liable for a range of different behaviors or activities into a (subjective) probability function.").

[11] *See, e.g.*, Cooper Indus., Inc. v. Leatherman Tool Grp., Inc., 532 U.S. 424, 438–39 (2001) (observing that punitive damages may be imposed in part to offset insufficient deterrence); United States v. Rogan, 517 F.3d 449, 454 (7th Cir. 2008) ("The lower the rate of a fraud's detection, the higher the multiplier required to ensure that crime does not pay." (citing Polinsky & Shavell, *Economic Analysis*, *supra* note 4)); United States v. Elliott, 467 F.3d 688, 692–93 (7th Cir. 2006) (using mathematical probability to determine an offender's expected gain from the crime); Parks v. Wells Fargo Home Mortg., Inc., 398 F.3d 937, 943 (7th Cir. 2005) ("One of the purposes of punitive damages is to punish a defendant who might otherwise find that its behavior was cost-effective." (citing Polinsky & Shavell, *Economic Analysis*, *supra* note 4, at 887)).

[12] As Itzhak Gilboa recently observed, "One cannot help wondering if the lack of concrete empirical testing in much of economic theory may have helped a beautiful but unrealistic [mathematical probability] paradigm to dominate the field." Itzhak Gilboa, *Questions in Decision Theory*, 2 ANN. REV. ECON. 1, 6 (2010).

[13] *See* Amos Tversky & Daniel Kahneman, *Evidential Impact of Base Rates*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 153, 154–60 (Daniel Kahneman et al. eds., 1982).

[14] This cognitive phenomenon is associated with overconfidence. *See* Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CALIF. L. REV. 1051, 1091–95 (2000); *see also* Sean Hannon Williams, *Sticky Expectations: Responses to Persistent Over-Optimism in Marriage, Employment Contracts, and Credit Card Use*,

probabilities of familiar scenarios,[15] and commit various other errors in carrying out probabilistic calculus.[16] These claims are typically substantiated by reference to experimental studies that identify people's probabilistic errors as systematic rather than accidental.[17] Scholars who challenge the empirical validity of the axiomatized view recommend that policymakers set up regulations that will keep people's risky choices on what they consider to be the right track. This track, so goes the argument, was paved by the mathematical system of probability.[18] These scholars also propose their own methods of improving the semi-rational probability assessments of an ordinary person. Those methods include de-biasing and other manipulations that make ordinary people's decisions correspond to the axiomatized view.[19]

This Article takes a fundamentally different route. In the pages ahead, I question the normative credentials of the axiomatized view and criticize its unreflective endorsement by lawyers, economists, and psychologists. The axiomatized view has established its dominance through the suppression and systematic neglect of an alternative system of probability: the "causative system." As I explain below, the causative system of probability allows people to make decisions compatible with the causal structure of their physical, social, and legal environments. Correspondingly, this system understands probability as a qualitative concept rather than a quantitative one.

---

84 NOTRE DAME L. REV. 733 (2009) (providing a general overview of the impact of overconfidence on decisionmaking).

[15] *See* Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, *supra* note 13, at 163, 163–64; Amos Tversky & Daniel Kahneman, *Judgments of and by Representativeness*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, *supra* note 13, at 84, 91–96.

[16] *See, e.g.*, Ehud Guttel, *Overcorrection*, 93 GEO. L.J. 241 (2004) (reporting that people overvalue refutations); Ehud Guttel & Alon Harel, *Matching Probabilities: The Behavioral Law and Economics of Repeated Behavior*, 72 U. CHI. L. REV. 1197 (2005) (reporting that people who engage in repeated activity match probabilities instead of relying on one correct probability estimate); Korobkin & Ulen, *supra* note 14, at 1085–95 (explaining the effects of availability, representativeness, and self-serving biases); Cass R. Sunstein, *Probability Neglect: Emotions, Worst Cases, and Law*, 112 YALE L.J. 61 (2002) (explaining people's tendency to focus on adverse outcomes rather than probabilities when intense emotions are involved). *See generally* Christine Jolls et al., *A Behavioral Approach to Law and Economics*, 50 STAN. L. REV. 1471 (1998) (providing an excellent overview of behavioral law and economics).

[17] *See* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, *supra* note 13, at 23–100, 153–208 (assembling relevant experimental studies).

[18] *See generally* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS (2008) (arguing that people systematically misjudge probabilities and make other cognitive errors and that government should set up a system of "choice architecture" that will nudge people's decisions in the right direction); Richard H. Thaler & Cass R. Sunstein, *Libertarian Paternalism*, 93 AM. ECON. REV. 175 (2003) (same). For excellent overview and discussion of these regulatory proposals, see Jonathan Klick & Gregory Mitchell, *Government Regulation of Irrationality: Moral and Cognitive Hazards*, 90 MINN. L. REV. 1620 (2006).

[19] *See, e.g.*, Baruch Fischhoff, *Debiasing*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, *supra* note 13, at 422, 423–27 (outlining debiasing procedures).

Causative probability originates from the writings of John Stuart Mill[20] and Francis Bacon.[21] Based on these philosophers' insights, it rejects the association of probability with instantial multiplicities.[22] The causative system uses a completely different criterion for ascribing probabilities to uncertain events: "evidential variety." This qualitative criterion focuses on the proximity of individuated causal scenarios as an empirical matter, and this proximity, in turn, depends on the wealth of confirmatory evidence. Confirmatory evidence denotes the presence of factors that tend to bring about the event in question and the presence of factors that negate rival causal scenarios or hypotheses. The number and variety of an event's evidentiary confirmations determine its causative probability. Most important, the reasoner's assessment of this probability ought to be case-specific and strictly empirical: the reasoner ought to ascribe no probative value whatsoever to purely statistical possibilities that her case-specific evidence does not confirm. This feature separates the causative system of probability from the mathematical system.[23]

The two systems of probability assessments not only are logically distinct from each other but also, more often than not, yield dramatically different results. Consider the following illustration:

> Peter undergoes a brain scan by MRI, and the scan is analyzed by a radiologist. The radiologist tells Peter that the lump that appears on the scan is benign to the best of her knowledge. She clarifies that she visually examined every part of Peter's brain and found no signs of malignancy. Peter asks the radiologist to translate the "best of her knowledge" into numbers, and the radiologist explains that 90% of the patients with similar-looking lumps have no cancer and that indications of malignancy are accidentally missed in 10% of the cases. The radiologist also tells Peter that only complicated brain surgery and a biopsy can determine with certainty whether he actually has cancer. According to the radiologist, this surgery involves a 15% risk of severe brain damage; in the remaining 85% of the cases, it successfully removes the lump and the patient recovers. Peter's primary care physician subsequently informs him that MRI machines have varying dependability. Specifically, he tells Peter that about 10% of those machines fail to reproduce images of small-size malignancies in the brain.

Under the mathematical system, Peter's probability of not having cancer equals 0.81. This number aggregates two probabilities of 0.9: the probability of correctness that attaches to the radiologist's diagnosis and the

---

[20] *See* JOHN STUART MILL, A SYSTEM OF LOGIC, RATIOCINATIVE AND INDUCTIVE: BEING A CONNECTED VIEW OF THE PRINCIPLES OF EVIDENCE AND THE METHODS OF SCIENTIFIC INVESTIGATION (photo. reprint 1980) (New York, Harper, 8th ed. 1874) (1843).

[21] *See* FRANCIS BACON, NOVUM ORGANUM (1620), *reprinted in* BACON'S NOVUM ORGANUM 191 (Thomas Fowler ed., Oxford, Clarendon Press, 2d ed. 1889).

[22] *See* COHEN, *supra* note 5, at 145–56.

[23] *See* L. JONATHAN COHEN, THE PROBABLE AND THE PROVABLE 34 (1977) (explaining mathematical probability as a complete system in which any scenario is considered probable until evidence affirmatively rules it out).

machine's probability of properly reproducing the image of Peter's brain.Peter's probability of having cancer consequently equals 0.19 (1 – 0.81).[24] This probability is greater than the 0.15 probability of sustaining severe brain damage from the surgery. Should Peter opt for the surgery?

Under the mathematical system, he should. The fatalities to which the two probabilities attach are roughly identical. If so, Peter should choose the course of action that reduces the fatality's probability. Under the mathematical system of probability, this choice will improve Peter's welfare (by 4% of the value of his undamaged brain).

Bacon and Mill, however, would advise Peter to rely on the causative probability instead. Specifically, they would tell Peter to rely on the radiologist's negative diagnosis and pay little or no attention to the background statistics. The radiologist's diagnosis is the only empirically-based causative account that concerns Peter's brain *individually*.[25] It identifies benignancy indications that appeared in that specific brain. The radiologist's reliance on those indications satisfies the evidential variety criterion.[26] As such, it is epistemically superior to the information about her and the MRI machine's general rate of error. Most crucially, the radiologist's diagnosis is the only evidence compatible with the causal nature of Peter's physical environment. The general statistic extrapolated from the radiologist's and the MRI machine's history of errors is fundamentally incompatible with this environment. This statistic identifies no causal factors that could foil the radiologist's diagnosis of Peter's brain.

Bacon and Mill would be right. Peter, indeed, should rely on the radiologist's diagnosis. He willmake a serious and potentially fatal mistake if he chooses to undergo the brain surgery instead. Evidence that the radiologist erred in the past in 10 diagnoses out of 100 reduces the general reliability of her diagnoses. This evidence, however, is causally irrelevant to the question of whether Peter has cancer. Whether Peter has cancer is a matter of empirical fact that the radiologist tried to ascertain. Her ascertainment of this fact relied on a series of patient-specific observations and medical science. The radiologist did not proceed stochastically by randomly distributing ten false-negative diagnoses across one hundred patients. Rather, she did her best for each and every patient, but, unfortunately, failed to

[24] This calculation applies the "negation rule." *See infra* Part I.A. The same probability can be calculated by aggregating Peter's 10% chance of having a small malignancy missed by the MRI machine with his 10% chance of being one of the radiologist's false negatives. Peter's probability of falling into *either* of these misfortunes equals (0.1 + 0.1) − (0.1 × 0.1) = 0.19. This calculation follows the "disjunction rule." *See infra* Part I.A.

[25] *Cf.* L. Jonathan Cohen, *Bayesianism Versus Baconianism in the Evaluation of Medical Diagnoses*, 31 BRIT. J. PHIL. SCI. 45 (1980) (arguing that patient-specific diagnoses are superior to statistical ones).

[26] *See, e.g.*, Michael Mavroforakis et al., *Significance Analysis of Qualitative Mammographic Features, Using Linear Classifiers, Neural Networks and Support Vector Machines*, 54 EUR. J. RADIOLOGY 80 (2005) (specifying malignancy and benignancy indicators that a radiologist should evaluate qualitatively in each patient and developing a quantitative tool to make those evaluations more robust).

identify cancer in 10 patients out of 100. These errors had patient-specific or scan-specific causes: invisible malignancies, malfunctioning MRI machines, accidental oversights, and so forth. Those causes are unidentifiable, which means that Peter may be among the afflicted patients. As an empirical matter, however, the unknown status of those causes does not equalize the chances of being misdiagnosed for each and every patient. Peter therefore has no empirical basis to discount the credibility of the radiologist's diagnosis of his brain by 10%.[27] This diagnosis is not completely certain, but it is supported by a solid causal theory: the radiologist's application of medical science to what she saw in Peter's brain. On the other hand, no causal theory can establish that the radiologist's patients are equally likely to be misdiagnosed as cancer-free.[28]

The epistemic virtue of causative probability has far-reaching implications for law enforcement. Law enforcement is an inherently causal activity. Courts, prosecutors, and other law enforcers do not define their tasks by throwing a die or by flipping a coin. Their implementation of legal rules is triggered by the evidence as to what the relevant actor did or did not do. Legal rules that law enforcers implement are causative as well. Virtually all of the rules focus on people's actions and the actions' consequences. Those rules set up mechanisms that allow people to reap the benefits of their productive activities and force them to pay for the harms they cause. All this turns causative probability into a primary tool for understanding how law enforcement mechanisms work and for improving the functioning of those mechanisms. This probability can both explain and guide lawenforcement decisions better than the mathematical system. Causative probability is also a superior tool for understanding the formation of individuals' incentives to comply with legal rules.

Policy recommendations that evolve from this insight cut across mainstream economic theory and behavioral economics. I argue that mainstream economic theory ought to revise all of its reform proposals that rely upon general mathematical probability. Specifically, I criticize two central tenets of law and economics: the argument that high penalties can compensate for a low probability of enforcement[29] and the assumption that individuals base their choices of action on the background statistics of accidents and harm.[30] My analysis also calls for a thorough revision of the behavioral theory that

[27] The same holds true for a possible malfunctioning of the MRI machine that scanned Peter's brain. There is no reason to believe that the risk of malfunction is distributed evenly across all machines and patients.

[28] Error statistics are not immaterial: if many (say, 30%) of the radiologist's diagnoses were false, Peter would have a good reason to doubt her credibility. This factor, however, would still be causatively irrelevant to whether he actually has cancer. Under these circumstances, Peter would have to find a credible specialist or endure the uncertainty. *Cf.* Judith Jarvis Thomson, *Remarks on Causation and Liability*, 13 PHIL. & PUB. AFF. 101, 127–33 (1984) (distinguishing between "external" evidence that derives from naked statistics and "internal" case-specific evidence that fits into a causal generalization).

[29] *See infra* Part II.A.

[30] *See infra* Part II.B–C.

diagnoses systematic failures in individuals' calculations of mathematical probability. Behavioral economists take mathematical probability as the benchmark for their appraisals of individuals' rationality without acknowledging the presence and viable functioning of the causative system. Indeed, I demonstrate that participants in core behavioral experiments executed their tasks in accordance with the causative probability system. They preferred to base their decisions on evidential variety, while paying little or no attention to instantial multiplicities.[31]

Structurally, the Article proceeds as follows: In Part I, I outline the principles of mathematical probability and identify their epistemic distortions. In Part II, I illustrate these distortions through economically driven court decisions and analyses of legal doctrines. In Part III, I explain how the causative system of probability works and show how it eliminates the distortions identified in Parts I and II. Part IV details my policy recommendations. Chief among those is a comprehensive shift from the mathematical to the causative probability system in the formation of legal incentives. A short conclusion follows.

I. MATHEMATICAL PROBABILITY: LANGUAGE AND EPISTEMICS

The best way to understand mathematical probability is to perceive it as a language that describes the facts relevant to a person's decision. Like all languages that people use in their daily interactions, the probability language has a set of conventional rules. These rules determine the meanings, the grammar, and the syntax of probabilistic propositions. Compliance with these rules enables one person to form meaningful propositions about probability and communicate them to other people.

The probability language differs from ordinary languages in three fundamental respects: scope, parsimony, and abstraction. First, ordinary languages have a virtually unlimited scope, as they promote multiple purposes in a wide variety of ways. People use those languages in communicating facts, thoughts, ideas, feelings, emotions, sensations, and much else. The probability language, in contrast, has a much narrower scope because it only communicates the reasoner's epistemic situation or balance of knowledge versus ignorance. The reasoner uses this language to communicate what facts she considers relevant to her decision and the extent to which those facts are probable. Second, ordinary languages have rich vocabularies.[32] The probability language, by contrast, is parsimonious by design: it uses a small set of concepts to describe multifarious events in a standardized mode. This mode establishes a common metric for all propositions about the probabilities of uncertain events. This metric creates syntactical uniformity in the probability language and makes it interpersonally transmit-

---

31 *See infra* Part IV.B.

32 *See, e.g.*, THE OXFORD ENGLISH DICTIONARY (2d ed. 1989) (a twenty-volume dictionary that explains the meanings of over 600,000 words originating from approximately 220,000 etymological roots).

table. Finally, because a person usually needs to deal with more than one uncertain event, she needs a uniform set of abstract concepts by which to relate one probability estimate to another and to integrate those estimates into a comprehensive assessment of probability.

These attributes of the probability language account for its high level of abstraction, uncharacteristic of any ordinary language. To maintain the required parsimony and conceptual uniformity, the probability language uses mathematical symbols instead of words. Those symbols allow a person to formulate her assessments of probability with precision. This precision, however, is purchased at a steep price: the comprehensive trimming of particularities and nuances that characterize real-world facts. The scope of each assessment's meaning and applicability thus becomes opaque and at times indeterminable (as illustrated by my introductory example of a bewildered patient who tries to figure out what medical statistics actually say about his condition). This tradeoff—precise language fora weak epistemic grasp—is a core characteristic and the core problem of mathematical probability. The two components of this tradeoff stand in an inverse relationship to each other. To be able to formulate her probability assessments with precision, a person must get rid of untidy concepts, downsize her vocabulary, and abstract away the multifaceted nuances of the real world. All this weakens the person's epistemic grasp of the world. As a result, her abstract, numerical estimates will say hardly anything informative about concrete events that unfold on the ground.

To have a strong epistemic grasp of the factual world, a person has to be wordy: she must utilize a rich vocabulary and loosen her conceptual precision. Indeed, as I show in Part III, the causative probability system does exactly this: it strengthens a decisionmaker's epistemic grasp at the expense of eroding the precision of her probability assessments. The benefits and the costs of this tradeoff are discussed in Part III as well. In this Part of the Article, I focus solely on the language and the epistemics of the mathematical system of probability.

A. Language: Using Numbers Instead of Words

The mathematical probability system designates the numerical space between 0 and 1 (the algebraic equivalents of 0% and 100%) to accommodate every factual scenario that exists in the world[33]:

0 ________________________________ 1

This space accommodates two propositions that are factually certain:

[33] My discussion simplifies Kolmogorov's classic definition of the "probability space." *See* A.N. KOLMOGOROV, FOUNDATIONS OF THE THEORY OF PROBABILITY (Nathan Morrison ed., 2d English ed. 1956).

PROPOSITION A: The probability that one of all the possible scenarios will materialize equals 1.

PROPOSITION B: Correspondingly, the probability that none of all the possible scenarios will materialize equals 0.

These propositions are tautological. The first proposition essentially says,"Something will certainly happen." The second makes an equally vacuous attestation: "There is no way that nothing will happen." All other propositions occupying the probability space are meaningful because they describe concrete events that unfold in the real world. These meaningful propositions are inherently uncertain. There is no way of obtaining complete information that will verify or refute what they say. Consequently, the probability of any concrete scenario is always greater than zero and less than one. More precisely, the probability of any concrete scenario,$P(S)$, equals one minus the probability of all factual contingencies in which the scenario does not materialize: $P(S)=1–P(\text{not-}S)$. This formula is called the "complementation principle."[34]

Here is a simple illustration of that principle:

0	0.5	1
$P(S)$		$P(\text{not-}S)$

Consider a random toss of a coin. The coin is unrigged: its probability of landing on heads is the same as itsprobability of landing on tails. Each of these probabilities thus equals 0.5. The two probabilities divide the entire probability space. The coin's probability of landing on either heads or tails equals 1, and we already know that this proposition is vacuous or tautological.

This illustration does not address the key question about the coin. What does "unrigged" mean? How does one know that *this* coin is equally likely to land on heads and on tails? This question is very important, but I intentionally do not address it here. This question focuses on the epistemic aspect of mathematical probability, discussed in section B below, while presently my only concern is the probability's syntax and semantics. For that reason, I simply assume that the two probabilities are equal.

We are now in a position to grasp the second canon of mathematical probability: the "multiplication principle" or the "product rule."[35] The multiplication principle holds that the probability of a joint occurrence of two mutually independent events,S_1 and S_2, equals the probability of one event multiplied by the probability of the other. Formally:

$$P(S_1\&S_2)=P(S_1)\times P(S_2)$$

---

[34] *See* COHEN, *supra* note 5, at 17–18, 56–57 (stating and explaining the complementation principle).

[35] *Id*. at 18–19 (stating and explaining the multiplication principle).

My coin example makes this principle easy to understand. Consider the probability of two successive tosses of an unrigged coin landing on heads. The probability that the first toss will produce heads, $P(S_1)$, equals 0.5. The probability that the second toss will produce heads, $P(S_2)$, equals 0.5 as well. The first probability occupies half of the entire probability space, while the second—as part of the compound, or conjunctive, scenario we are interested in—occupies half of the space taken by the first probability. The diagram below shows this division of the probability space:

$P(S_1 \& S_2)$
————|————|————————
0   0.25   0.5   1
$P(S_1)$

The complementation and multiplication principles are the pillars of the mathematical system of probability. All other probability rules derive from these principles. Consider the "disjunction rule"[36] that allows a person to calculate the probability of alternative scenarios, denoted again as S_1 and S_2. This probability equals the sum of the probabilities attaching to those scenarios, minus the probability of the scenarios' joint occurrence. Formally: $P(S_1 \text{ or } S_2)=P(S_1)+P(S_2) - P(S_1 \& S_2)$. Here, the deduction of the joint-occurrence probability, $P(S_1 \& S_2)$, prevents double counting of the same probability space. The probability of each individual scenario, $P(S_1)$ and $P(S_2)$, occupies the space in which the scenario unfolds both alone as well as in conjunction with the other scenario: $P(S_1)$ occupies the space in which S_1 occurs together with S_2, and $P(S_2)$ occupies the space in which S_2 occurs together with S_1. There is, however, only one space for $S_1 \& S_2$ as a combined scenario, and hence the deduction.

A joint occurrence of two (or more) events is not always factually possible. For example, a single toss of a coin can yield either heads or tails, but not both: that is, $P(S_1 \& S_2)=0$. The coin's probability of landing on heads or, alternatively, on tails consequently equals 1 (0.5+0.5 – 0). But in real-life situations, events often occur in conjunction with each other. For example, a medical patient's permanent disability may originate from his preexisting condition, his doctor's malpractice, or from both. If so, then $P(S_1 \& S_2) > 0$.

A conjunctive occurrence of two events can also be perceived as a compound scenario in which one event (H) unfolds in the presence of another (E). The probability of any such scenario is called "conditional" because it does not attach unconditionally to a single event, H, but rather to event H given the presence, or occurrence, of E, which is denoted as $P(H|E)$.

---

[36] *See* WILLIAM KNEALE, PROBABILITY AND INDUCTION 125–26 (1949) (stating and explaining the disjunction rule).

This formulation allows me to present the last basic component of the mathematical probability system: the Bayes Theorem.[37] This theorem establishes that when I know the individual probabilities of E and H and the probability of E's occurrence in the presence of H, I can calculate the probability of H's occurrence in the presence of E. Application of the multiplication principle to the prospect of a joint occurrence of two events, E and H, yields $P(E\&H)=P(E)\times P(H|E)$. Under the same principle, the conjunctive probability of E and H, restated as $P(H\&E)$, also equals $P(H) \times P(E|H)$. This inversion sets up a probabilistically important equality: $P(E) \times P(H|E) = P(H) \times P(E|H)$.[38] The Bayes Theorem is derived from this equality: $P(H|E) = P(H) \times P(E|H) \div P(E)$.

My labeling of the two events as E and H is not accidental. Under the widely accepted terminology, H stands for a reasoner's *hypothesis*, while E stands for her *evidence*. Both E and H are events, but the reasoner is not considering those events individually. Rather, she is examining the extent to which evidence E confirms hypothesis H. The Bayesian formulation consequently separates between the probability of hypothesis H before the arrival of the evidence ($P(H)$); the general probability of the evidence's presence in the world ($P(E)$); and the probability of the evidence being present in cases in which hypothesis H materializes ($P(E|H)$). These three factors allow the reasoner to compute the posterior probability of her hypothesis: the probability of hypothesis H given evidence E.

The reasoner must process every item of her evidence sequentially by applying this procedure. She must perform the Bayesian calculation time and time again until all of her evidence is taken into account. Each of those calculations will update the hypothesis's prior probability by transforming it into a new ("posterior") probability. The posterior probability will become final after the reasoner had exhausted all of the available evidence.[39]

Consider the significance of the evidence-based multiplier, $P(E|H) \div P(E)$. This multiplier is called the "likelihood ratio"[40] or—as I prefer to call it—the "relevancy coefficient."[41] The relevancy coefficient measures the frequency with which E appears in cases featuring H, relative to the frequency of E's appearance in all possible cases. If $P(E|H) \div P(E) > 1$ (E's appearance in cases of H is more frequent than its general appearance), the probability of hypothesis H goes up. Formally: $P(H|E) > P(H)$, which means that evidence E confirms hypothesis H. If

[37] *See* Thomas Bayes, *An Essay Towards Solving a Problem in the Doctrine of Chances*, *in* PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY OF LONDON (1763), *available at* http://www.stat.ucla.edu/history/essay.pdf (last visited Nov. 20, 2010). For a modern statement of the theorem, see COHEN, *supra* note 5, at 68.

[38] Because of this inversion, some call the Bayes Theorem the "Inversion Theorem." *See, e.g.*, KNEALE, *supra* note 36, at 129.

[39] For a good explanation of this updating, see DAVID A. SCHUM, THE EVIDENTIAL FOUNDATIONS OF PROBABILISTIC REASONING 215–22 (1994).

[40] *Id.* at 218.

[41] *Id.* at 219 (associating the likelihood ratio with the "force of evidence").

$P(E|H) \div P(E) < 1$ (E's appearance in cases of H is less frequent than its general appearance), the probability of hypothesis H goes down. Formally: $P(H|E) < P(H)$, which means that evidence E makes hypothesis H less probable. Finally, if $P(E|H) = P(E)$ (E's appearance in cases of H is as frequent as its general appearance), the presence of E does not influence the probability of H. This makes evidence E irrelevant.[42]

To illustrate, consider a tax agency that uses internal fraud-risk criteria for auditing firms.[43] By applying those criteria, the agency singles out for auditing one firm out of ten. This ratio is public knowledge. Firms do not know anything about the agency's criteria for auditing (nor does anyone else outside the agency). Under this set of facts, each firm's prior probability of being audited equals 0.1.

Now consider an individual firm whose reported expenses have doubled relative to past years. Does this evidence change the probability of being audited? The answer to this question depends on whether a steep increase in a firm's reported expenses appears more frequently in cases in which it was audited than in general. Assume that experienced accountants formed an opinion that increased expenses are three times more likely to appear in auditing situations than generally. This relevancy coefficient triples the prior probability of the firm's audit. The firm's posterior probability of being audited thus turns into 0.3.

But how do we know that these evidential effects are brought about by causes and more than a mere correlation? We do not know it for sure, and I address this issue in section B. Here, I focus on the semantics and syntax of mathematical probability. The Bayes Theorem is part of those semantics and syntax: it tells us how to conceptualize our epistemic situations by using mathematical language. As the following discussion demonstrates, the theorem itself provides no instructions on how to grasp causes and effects of the outside world and relate them to each other.

Mathematical language creates a uniform conceptual framework for all probability assessments that rely on instantial multiplicity. For those who base their estimates of probability on events' frequency, this language is indispensable.[44] This language is also necessary for formulating probability assessments on the basis of propensity—a disposition of a given factual setup to produce a particular outcome over a series of cases or experiments.[45]

---

[42] *Cf.* Richard O. Lempert, *Modeling Relevance*, 75 MICH. L. REV. 1021, 1025–27 (1977) (offering similar formulation of relevancy coefficients).

[43] A good real-world example of this practice is the secret "Discriminant Index Function" (DIF), used by the IRS in selecting taxpayers for audits. *See, e.g.*, Gillin v. Internal Revenue Serv., 980 F.2d 819, 822 (1st Cir. 1992) ("The IRS closely guards information concerning its DIF scoring methodology because knowledge of the technique would enable an unscrupulous taxpayer to manipulate his return to obtain a lower DIF score and reduce the probability of an audit."); Sarah B. Lawsky, *Probably? Understanding Tax Law's Uncertainty*, 157 U. PA. L. REV. 1017, 1068–70 (2009) (describing the DIF method used by the IRS).

[44] *See* COHEN, *supra* note 5, at 47–48 (explaining frequency as a rate of relevant instances).

[45] *Id*. at 53–58 (explaining propensity as a rate of relevant instances).

Finally, people basing their decisions upon intuited or "subjective" probabilities[46] must use mathematical language as well. This language introduces conceptual precision and coherence into a reasoner's conversion of her experience-based beliefs into numbers. Those numbers must more or less correspond to the reasoner's empirical situation. A mismatch between the numbers and empirical reality will produce a bad decision.[47]

Proper use of the mathematical language does not guarantee that a person's probability assessments will be accurate. Mathematical language only helps a person conceptualize her raw information in numerical terms and communicate it to other people. Before using this language, a person must properly perceive and understand this information. This basic cognitional task is antecedent to a person's mathematical assessment of probability.

Proper use of the mathematical probability system thus can only guarantee a particular kind of accuracy: accuracy in ascribing probability estimates to perceived generalities, as opposed to individual events. Assuming that a person is able to properly conceptualize her experiences in mathematical language, will her probability assessments be accurate if she commits no mathematical errors in making those assessments? This question is fundamental to the entire probability theory, and the answer to it depends on what "accurate" means. The mathematical system offers no event-specific guarantees of accuracy in probability assessments. As the famous saying goes, for statistics there are no individuals, and for individuals no statistics.[48]

Accuracy is an inherently relational concept: its meaning cannot be determined independently of the phenomenon that a person attempts to understand and explain. The criteria for accuracy partly derive from the nature of that phenomenon. Consider a person who attempts to make sense of some general regularity that exists in the world: say, "The tax agency audits predominantly those firms that report high operational expenses." The accuracy of what the person comes to believe about this regularity ought to be general or statistical rather than inferred from a single event. The person cannot rationally believe in this regularity after being informed, for example, that tax officials have decided to audit a single firm that reported high operational expenses. To be able to confirm or disconfirm this regularity, the person must ascertain the number of audit instances involving firms with unusually high reported expenses in the general pool of audits. Alternatively, the person may slightly relax her accuracy criteria and rely on a large sample of tax-reporting firms. If that sample reveals a high proportion of audited firms with reportedly high operational expenses, it willconfirm the regularity.

---

[46] *Id*. at 58–70 (explaining subjective probability in terms of reasoners' betting odds).

[47] *See id*. at 60.

[48] *See, e.g.*, George O'Brien, *Economic Relativity*, 17 J. STAT. & SOC. INQUIRY SOC'Y IR. 1, 11 (1942) ("[F]or individuals there are no statistics, and for statistics there are no individuals.").

What accuracy criteria are appropriate for event-specific predictions? Consider a CEO attempting to find out whether *herfirm* will be audited after reporting high operational expenses. The CEO can hardly satisfy herself with the general rate of audits across firms. For her individual firm, this rate is empirically meaningless and uninformative. Whether this firm will actually be audited will be decided by individual causal factors, not by a random lottery.[49] The auditing decision will reflect the tax agency's reaction to the firm's audit-triggering activities. The CEO therefore should try to obtain information that willallow her to make an individuated causative assessment of the firm's prospect of being audited. As part of that inquiry, she ought to find out whether the firm's reported expense is associated with activities unquestionably related to its business. The CEO should also ascertain whether the firm was ever red-flagged by the tax agency. And she must consider other specifics of the firm's situation as well: for example, whether an employee recently fired by the firm had delivered on his threat to tell tax officials about the firm's accounting irregularities. In other words, the CEO should base her prediction upon the evidential variety that pertains to her case, as opposed to instantial multiplicity that pertains to all cases at once. This case-specific evidence may eliminate or, alternatively, affirm the presence of circumstances prompting the agency to audit the CEO's firm (as opposed to firms generally or an "average firm" with high reported expenses).

But what if the general distribution of audits werethe only evidence available to the CEO? In that scenario, the CEO might base her event-specific prediction on the general distribution. Doing so would not be irrational. However, the accuracy of the CEO's prediction would then be compromised. The CEO would hardly be able to recommend any specific action on the basis of this statistical prediction. For example, she would not hire an expensive accountant to carry out self-audit, nor would she commission an internal investigation of the firm's affairs. This low level of accuracy starkly contrasts with the high accuracy level characterizing the probabilistic assessment of the auditing regularity as a general proposition. For example, tax policy analysts who rely on this assessment have a strong epistemic grasp on the proposition "The tax agency audits predominantly those firms that report high operational expenses." The CEO, in contrast, would only have a weak epistemic grasp on the proposition "My firm will likely be audited because its reported expenses are high."

The difference between these epistemic grasps is fully explained in section B. For now, the readers only need to acknowledge its existence and intuitive appeal. With this in mind, I continue my tax-audit example. Assume that the CEO obtains the case-specific evidence for which she was looking. She learns from this evidence that her firm has never been red-

[49] If the agency were to audit one firm out of ten after selecting its auditees by a random draw in which all firms participated, the 0.1 probability of being audited would then be empirically significant.

flagged, that its reported expenses are unquestionably business-related, and that the employee it laid off has found a better job and is no longer resentful. Based on this information and on what she knows about the general distribution of audits across firms, the CEO now makes an assessment of her firm's probability of being audited. She attempts to derive this individuated causative assessment from the evidence that supports the audit and the no-audit scenarios *for her firm*. To this end, the CEO tries to utilize the mathematical language. Can she use this language to convert the evidence upon which she relies into a numerical estimate of probability?

To succeed in this task, the CEO first needs to articulate her best prediction. This articulation is easy to make: the existing evidence strongly (albeit not unequivocally) supports the prediction that the firm will not be audited. The CEO subsequently needs to position that evidence in the probability space, as required by the mathematical language. This positioning turns out to be a rather daunting task. Because the probability that the firm will either be audited or not equals 1 (recall the coin-flip example), any space between 0 and 1 not occupied by the audit-free scenarios is occupied by the scenarios in which the firm is audited. This mathematical rule overrides the empirical absence of audit evidence. The complementation principle deems this evidence to be present somewhere and somehow, unbeknownst to the CEO. Given the incompleteness of the CEO's evidence and what she knows about the distribution of audits across firms in general, this assumption is not completely unwarranted. However, what is unwarranted here is the numerical figure that purports to estimate the strength or significance of the audit evidence. Under the complementation principle, the strength of this completely unknown evidence equals 1 minus the strength of the known evidence that supports the no-audit prediction. This formula makes no epistemic sense at all because the unknown evidence could actually further confirm the CEO's existing evidence and make her prediction unassailable. There are no epistemic grounds upon which to assign the unoccupied probability space to evidence that is not present and that may actually not exist. The unoccupied space also cannot be allotted to the general statistical probability of *one* firm's audit. Evidence upon which the CEO bases her no-audit prediction excludes her firm from the statistical regularity that this probability represents. This regularity may somewhat weaken the overall strength of the CEO's case-specific evidence, and the CEO therefore needs to take it into account. The CEO, however, need not carry the mathematical figure representing this statistical regularity over to her case. This figure has no bearing on the CEO's individual case.

The multiplication principle would distort the CEO's assessment of the evidence equally badly. Under this principle, the CEO would have to multiply the quantified supports of her no-audit prediction by each other. This multiplication is alien to the CEO's epistemic endeavor and would produce an anomalous result. The CEO's task is not to calculate the ex ante chances of her evidential items' conjunctive occurrence. Rather, she tries to ascertain the ex post evidential effect of that occurrence on the individual causative scenario: the tax agency's reaction to the firm's reported expenses. To advance this inquiry, the CEO should evaluate the extent to which this evidential occurrence supports her no-audit prediction for the firm. Specifically, she needs to figure out whether her firm's circumstances and their evidentiary coverage eliminate the reasons prompting tax audits. Those reasons are general, but the firm's evidence is individual, and so is its effect on the agency's auditing decision. The CEO's evaluation of this effect will therefore be case-specific rather than statistical. She will try to form the best prediction with respect to her firm's audit on the basis of her evidence rather than calculate the percentage of cases in which predictions similar to hers come true.

The metric set by the mathematical language consequently does not help the CEO. This metric treats probability as coextensive with instantial multiplicity and recognizes no other criteria for probabilistic appraisals. As a result of this definitional constraint, the metric contains no quantifiers for evidential variety or for the degrees of evidential support for event-specific hypotheses. This limitation is profound: it makes mathematical language unfit as a tool for event-specific assessments of probability. Event-specific probabilities are conceptually nonmathematical.

One may respond that the CEO may still reconceptualize her evidence so as to make it fit the mathematical language. This reconceptualization is not difficult to carry out. The general rate of audits among firms with high reported expenses gives the CEO the prior probability with which to begin her inquiry. The CEO therefore needs to find out, or intuit, this rate. Subsequently, she ought to multiply this rate by the relevancy coefficient, as mandated by the Bayes Theorem. The CEO can determine this coefficient in three steps. First, she needs to determine, or intuit, the general recurrence rate for evidence similar to hers. Subsequently, she needs to determine, or intuit, the recurrence rate for having such evidence present in cases in which the firm is selected for audit. Finally, the CEO must divide the second number by the first.

To illustrate, assume that the tax agency audits fiveout of every tenfirms that report high operational expenses. Also assume that sevenout of every tenfirms with high reported expenses have expenses that are clearly business-related. Moreover, none of these firms is red-flagged or faces bad-accounting accusations from a former employee. Finally, assume that among every tenfirms audited by the agency only one exhibits these three characteristics at once. These figures yield a very low relevancy coefficient: 1/7.[50] The posterior probability of the firm's selection for audit con-

[50] For clarity's sake, I explain the calculation. The probability of the firm's type of evidence being present in the event of an audit is 1/10. The general probability of such evidence being present (in both audit and audit-free events) is 7/10. The relevancy coefficient represents the fraction of cases featuring high reported expenses, innocent explanation, and audit in the more general pool of cases that exhibit

sequently amounts to 0.07.[51] The firm's probability of not being audited that the CEO needs to ascertain equals 0.93 (1 – 0.07). This outcome is intuitive, and it also seems to correspond to the specifics of the case. If so, why not prefer this form of mathematical reasoning over that based upon evidential variety?

The answer to this question depends on two factors. The first factor is the nature of the occurrence that a person needs to evaluate in probabilistic terms. This occurrence can be a discrete, empirically verifiable event (e.g.,"The tax agency will audit my firm in the coming months"), or it can be a multiplicity of events presented as a generalization (e.g.,"The tax agency audits firms whose reported expenses are unusually high."). The second factor is the quality, or the strength, of the person's epistemic grasp of the occurrence. Mathematical language allows people to develop a strong epistemic grasp on abstractly formulated generalizations. This language, however, is not suitable for a person trying to establish a strong epistemic grasp of a single real-world event. The causative probability system and its evidential-variety criterion will serve that person better.

In some cases, as in my tax-audit example, a person may arrive at similar assessments of probability under both systems. This similarity, however, is merely coincidental. There is no guarantee that it will be present in every case or even in the majority of the cases. As I demonstrated in the Introduction, the two systems may give people conflicting recommendations on matters of life and death. My argument that causative probability improves a person's epistemic grasp of individual events—relative to the grasp she can achieve under the mathematical system—therefore has far-reaching implications in both practical and theoretical domains.

Thisalleged improvement is in need of further articulation. Thus far, I have established its presence on the conceptual level and, hopefully, on the intuitive level as well. I have yet to carry out a rigorous epistemological comparison between the two systems of probabilistic reasoning. This comparison is crucial. The fact that the causative system aligns with common sense; comports with the causal structure of people's physical, social and legal environments; and offers a convenient taxonomy for probabilistic assessments of individual events speaks in that system's favor. This fact, however, is not decisive. In order to establish the superiority of the causative system, one also needs to show that its rules of inference outscore those of the mathematical system in the domain of epistemology. One needs to demonstrate, in other words, that the causative probability rules improve the accuracy of people's probabilistic assessments of individual events. In the remainder of this Article, I attend to this task.

---

high reported expenses and innocent explanation, both with and without an audit. This coefficient equals 1/10 ÷ 7/10, i.e., 1/7 or 0.14.

[51] This probability equals the firm's prior probability of being audited (0.5) multiplied by the relevancy coefficient (0.14).

B. *Epistemics: Instantial Multiplicity as a Basis for Inference*

John Stuart Mill sharply criticized the use of instantial multiplicity as a basis for inference.[52] He described it as "the natural Induction of uninquiring minds, the induction of the ancients, which proceeds *per enumerationem simplicem*: 'This, that, and the other A are B, I cannot think of any A which is not B, therefore every A is B.'"[53]

This sentence succinctly identifies the epistemological weakness of the mathematical probability system. The system's mathematical rules instruct the reasoner on how to convert her information into cardinal numbers. These rules have no epistemic ambition. They do not tell the reasoner what counts as information upon which she ought to rely. This task is undertaken by the system's rules of inference. I examine those rules in the paragraphs ahead.

One of those rules holds that any scenario not completely eliminated by existing evidence is a factual possibility that must occupy some of the probability space. The reasoner must consequently assign some probability to any such scenario, and this probability must be greater than zero. I call this rule "the uncertainty principle."

The second rule—"the principle of indifference"—is a direct consequence of the first. This rule determines the epistemic implications of the unavailable information for the reasoner's probability decision. The rule postulates that unavailable information is not slanted in any direction, meaning that the reasoner has no reasons for considering one unevidenced scenario as more probable than another unevidenced scenario.[54] The reasoner ought to be epistemically indifferent between those scenarios, and this indifference makes the unevidenced scenarios equally probable.

The third rule logically derives from the second. It presumes that statistical distributions are extendible. To follow Mill's formulation, if 70% of events exhibiting feature *A* exhibit feature *B* as well, then presumptively any future occurrence of *A* has a 70% chance of occurring together with *B*. I call this rule "the extendibility presumption." This presumption is tentative and defeasible: new information showing, for example, that *B* might be brought about by *C*—a causal factor unassociated with *A*—would render it inapplicable. Absent such information, however, the extendibility presumption applies with full force. The presumption's mechanism relies on the indifference principle as well. This principle treats all indistinguishable occurrences of *A*, past and future, as equivalents. The same principle marks any missing information that could identify *B*'s causal origins as unslanted. The reasoner consequently must treat this unknown information as equally likely to both increase and decrease the rate of *B*'s appearance in cases of *A*.

52 *See* MILL, *supra* note 20, at 549–53.
53 *Id*. at 549.
54 *See* COHEN, *supra* note 5, at 43–44.

Every future occurrence of A thus becomes statistically identical to A's past occurrences that exhibited B at a 70% rate.

The uncertainty principle seems epistemologically innocuous, but this appearance is misleading. Any factual scenario that existing evidence does not completely rule out must, indeed, be considered possible. This scenario therefore must have *some* probability on a 0–1 scale. All of this is undoubtedly correct. The uncertainty principle, however, also suggests that the reasoner *can* assign concrete probabilities to such unevidenced scenarios. This "can" is epistemologically unwarranted because the reasoner does not know those probabilities. Any of her probability estimates willbe pure guesswork: a creation of knowledge from ignorance.

To illustrate, consider an infinitesimally small, but still positive, probability that the Boston Red Sox will recruit me as a pitcher for next season. (There is no evidence that precludes this scenario completely.) Other law professors may have only slightly better probabilities of becoming Major League Baseball players. Each of these probabilities is close to, but still greater than, zero. Aggregation of these unevidenced probabilities might nonetheless yield a nonnegligible number. The probability of the scenario in which an MLB team drafts *a* law professor equals the sum of these probabilities, minus the probability of two or more professorial recruitments.[55] From a purely logical viewpoint, this number is unassailable: outside the realm of the impossible, any event has a chance to occur; and the more chances are present, the more likely is one of them to materialize. As an empirical matter, however, this number makes no sense at all.

The principle of indifference is a pillar of the entire system of mathematical probability.[56] It stabilizes the reasoner's information in order to make it amenable to mathematical calculus.[57] The principle's information-stabilizing method is best presented in Bayesian terms. Take a reasoner who considered all available information and determined the probability of the relevant scenario,$P(S)$. The reasoner knows that her information is incomplete and turns to estimating the implications of the unavailable information (U). The reasoner tries to figure out whether this unavailable information could change her initial probability estimate,$P(S)$. In formal terms, the reasoner needs to determine $P(S|U)$. Under the Bayes Theorem, which I have already explained, this probability equals $P(S)\times [P(U|S) \div P(U)]$. With the prior probability,$P(S)$, already known, the reasoner needs to determine the relevancy coefficient,$P(U|S)\div P(U)$. To this end, she needs to obtain two probabilities: the probability of U's ap-

[55] The subtraction of the overlapping probability is necessary to prevent double counting. *See supra* Part I.A.

[56] *See* JOHN MAYNARD KEYNES, A TREATISE ON PROBABILITY 41–42 (1921) (describing the indifference principle as essential for establishing equally probable possibilities—a preliminary condition for all mathematical assessments of probability).

[57] As Keynes explains, "In order that numerical measurement may be possible, we must be given a number of *equally* probable alternatives." *Id.* at 41.

pearance in general and the probability of U's appearance in cases of S. Because the reasoner has no information upon which to make this determination, the indifference principle tells her to assume that U is not slanted. That is, the reasoner must assume that U is equally likely to confirm and to disconfirm S: $P(U|S) = P(U)$. The relevancy coefficient consequently equals 1, and the reasoner's prior probability, $P(S)$, remains unchanged. The indifference principle essentially instructs the reasoner to deem missing information altogether irrelevant to her decision.

This instruction is epistemologically invalid. The reasoner can treat unavailable information as irrelevant to her decision only if she has no reason to believe that it might be relevant.[58] Whether those reasons are present or absent depends on the reasoner's known information. When this information indicates that the unavailable information might be relevant, $P(U|S)$ and $P(U)$ can no longer be considered equal to each other. The indifference principle consequently becomes inapplicable. On the other hand, when the known information indicates that the unavailable information is irrelevant to the reasoner's decision, something else happens. The known information establishes that $P(U|S)$ actually equals $P(U)$. The proven, as opposed to postulated, equality between $P(U|S)$ and $P(U)$ makes the indifference principle redundant. From the epistemological point of view, therefore, there are no circumstances under which this principle can ever become applicable.[59]

The indifference principle does not merely purport to manage unavailable information. Instead, it forces itself on the available information by requiring the reasoner to interpret that information in a particular way. Effectively, the principle instructs the reasoner to proceed on the assumption that all the facts necessary for her probability assessment are specified in the available information. This artificially created informational closure sharply contrasts with the causative system's criterion for probability assessments: the *actualextent* to which the available information specifies the facts necessary for the reasoner's decision.[60] The indifference principle ascribes arbitrary probative value to information that the causative system requires reasoners to measure.

The extendibility presumption is an equally problematic device. This presumption bypasses the question of causation, which makes it epistemologically deficient.[61] As Mill's quote suggests, an occurrence of feature B

---

[58] *See id.* at 55–56.

[59] *See* COHEN, *supra* note 5, at 45–46 (showing that the indifference principle is either circular or redundant); KEYNES, *supra* note 56, at 45–47 (demonstrating that the indifference principle is arbitrary and epistemologically unsustainable).

[60] *See* L. Jonathan Cohen, *On the Psychology of Prediction: Whose Is the Fallacy?*, 7 COGNITION 385, 389 (1979) ("Baconian [causative] probability-functions . . . grade probabilification . . . *by the extent to which* all relevant facts are specified in the evidence.").

[61] Another problem with extendibility is its dependence on a reference class—a statistical generalization that can be gerrymandered in numerous ways. *See* Ronald J. Allen & Michael S. Pardo, *The Problematic Value of Mathematical Models of Evidence*, 36 J. LEGAL STUD. 107, 111–14 (2007).

in numerous cases of *A* does not, by and of itself, establish that *B* might occur in a future case of *A*. Only evidence of causation can establish that this future occurrence is probable. This evidence needs to identify the causal forces bringing about the conjunctive occurrence of *A* and *B*. Identification of those forces needs to rely on a plausible causal theory demonstrating that *B*'s presence in cases of *A* is law-bound rather than accidental.[62] This demonstration involves proof that *B* is or tends to be uniformly present in cases of *A* for reasons that remain the same in all cases.[63] Those invariant reasons make the uniformity law-bound.[64] Their absence, in contrast, indicates that *B*'s presence in cases of *A* is possibly accidental. The observed uniformity consequently becomes nonextendible. Decisionmakers who choose to rely on this uniformity will either systematically err or arrive at correct probability assessments by sheer accident. They will never base those assessments upon knowledge.[65]

To illustrate, consider again the basic factual setup of my tax-audit example: the tax agency audits one firm out of ten. Assuming that no other information is available, will it be plausible to estimate that each firm's probability of being audited equals 0.1? This estimate's plausibility depends on whether the "one-to-ten" distribution is extendible. This distribution could be extendible if the agency were to make its audit decisions by some randomized procedure, such as a draw. This randomization would then give every firm an equal chance of being audited by the agency. The agency, however, does not select audited firms by a draw. Instead, it applies its secret fraud-risk criteria. This fact makes the observed distribution of audits nonextendible. Consequently, the 0.1 estimate of a firm's probability of being audited is completely implausible. Relying on it would be a serious mistake.[66]

To rebut this critique, adherents of mathematical probability might invoke the long-run argument, mistakenly (but commonly) grounded upon

---

62 *See* L. JONATHAN COHEN, THE DIALOGUE OF REASON: AN ANALYSIS OF ANALYTICAL PHILOSOPHY 177 (1986); *see generally* Marc Lange, *Lawlikeness*, 27 NOÛS 1 (1993) (defining law-bound regularities as separate from accidental events).

63 *See* COHEN, *supra* note 62, at 177–79.

64 *Id*. at 179.

65 For classic accounts of why accidentally true beliefs do not constitute knowledge, see Edmund L. Gettier, *Is Justified True Belief Knowledge?*, 23 ANALYSIS 121 (1963), which explains that accidentally acquired justification for a true belief is not knowledge; and Alvin I. Goldman, *A Causal Theory of Knowing*, 64 J. PHIL. 357 (1967) (attesting that a knower's true belief must be induced by the belief's truth). *See also* ROBERT NOZICK, THE NATURE OF RATIONALITY 64–100 (1993) (defining knowledge as a true belief supported by the knower's truth-tracking reasons).

66 Taxpayers' responses to an increase in the general probability of audit are difficult to measure. For one such attempt, see Joel Slemrod et al., *Taxpayer Response to an Increased Probability of Audit: Evidence from a Controlled Experiment in Minnesota*, 79 J. PUB. ECON. 455, 465 (2001), which finds that audit rates are positively correlated with reported income of low-income and middle-income taxpayers and are negatively correlated with reported income of high-income taxpayers.

Bernoulli's law of large numbers.[67] This argument concedes that the 0.1 estimate of a firm's probability of being audited is not a reliable predictor of any specific auditing event. The argument, however, holds that repeat-players—firms that file tax reports every year—should rely on this estimate because, at some point it will transform into a real audit. With some firms, it will happen sooner than with others, but eventually the agency will audit every firm.

This argument recommends that every person perceive her epistemic state of uncertainty as a physical experience of a series of stochastic events that can take her life in any direction. This recommendation fills every informational gap with God playing dice. However, neither God nor the tax agency will actually throw a die to identify firms that require an audit. Whether a particular firm will be audited will be determined by causal forces, namely, the tax officers who will apply the agency's fraud-risk criteria to what they know about each firm. Each firm therefore should rely on its best estimate of how those officers will evaluate its tax return. If, instead of relying on this estimate, a firm chooses to base its actions on the 10% chance of being audited, it willsooner or later find itself on the losing side.[68] This firm willeither take wasteful precautions against liability for tax evasion or expose itself to that liability by acting recklessly.[69]

---

[67] *See* JACOB BERNOULLI, THE ART OF CONJECTURING 315–40 (Edith Dudley Sylla trans., 2006) (1713). For a superb account of the law's intellectual history, see IAN HACKING, THE TAMING OF CHANCE 95–104 (1990).

[68] This point was famously made by P.A. Samuelson, *Risk and Uncertainty: A Fallacy of Large Numbers*, 98 SCIENTIA 108 (1963).

[69] To mitigate this problem, statisticians often use "confidence intervals." *See, e.g.*, THOMAS H. WONNACOTT & RONALD J. WONNACOTT, INTRODUCTORY STATISTICS 253–86 (5th ed. 1990). A confidence interval is essentially a second-order probability: an estimate of the chances that the reasoner's event-related (first-order) probability is accurate. Conventionally, those chances must not go below 95%—a confidence level that promises that the reasoner's estimate of the event-related probability will be accurate in 95 cases out of 100. *Id*. at 254–55. The reasoner must conceptualize her estimate of the event-related probability not as a fixed figure, but rather—more realistically—as an average probability deriving from a sample of probabilities attaching to factual setups similar to hers. The reasoner should expand her sample of setups by relying on her experience or by conducting a series of controlled observations. If she obtains a sufficiently large sample, the setups' probabilities will form a "normal" bell-shaped distribution curve. Subsequently, in order to obtain a 95% confidence level in her estimate of the probability, the reasoner must eliminate the curve's extremes and derive the estimate from the representative middle. Technically, she must shorten the distribution curve by trimming away 2.5% from each tail. This trimming will compress the reasoner's information and narrow the range of probabilities in her sample. The average probability calculated in this way will then have a high degree of accuracy. The chances that it will require revision in the future as a result of arrival of new information are relatively low. This feature will make the probability estimate resilient or, as some call it, robust or invariant. *See* JAMES LOGUE, PROJECTIVE PROBABILITY 78–95 (1995) (associating strength of probability estimates with resiliency); ROBERT NOZICK, INVARIANCES: THE STRUCTURE OF THE OBJECTIVE WORLD 17–19, 79–87 (2001) (associating strength of probability estimates with their invariance across cases).

The 95% confidence-interval requirement undeniably improves the quality of probabilistic assessments. The fact that those assessments stay invariant across many instances makes them dependable. *See* COHEN, *supra* note 5, at 118. This improvement, however, does not resolve the deep epistemological problem identified in this section. Resilience of a probability estimate only indicates that the esti-

II. PROBABILISTIC DISTORTIONS IN LAW AND ECONOMICS

In Part I, I demonstrated that mathematical probability is both conceptually and epistemically incompatible with case-specific inquiries into whether one individual occurrence will bring about another individual occurrence. The mathematical system associates probability with instantial multiplicity alone. Consequently, it provides no metric for assessing the probabilistic effect of the evidential varieties that characterize individual events. The system's conceptual tools thus fail to capture the probability of individual occurrences. Application of those tools weakens the reasoner's epistemic grasp of those occurrences instead of improving it.

Law and economics scholars fail to recognize this fundamental incompatibility. Correspondingly, they fail to realize that mathematical probabilities exert virtually no influence on the formation of individuals' reasons for action. My goal here is to evaluate the extent and the consequences of this neglect without delving into its causes.[70] This neglect accounts for a num-

---

mate is statistically stable. For example, a resilient probability of 0.7 can only identify the number of cases—70 out of 100—in which the underlying event will actually occur. This assurance, however, does not determine the applicability of the 0.7 probability to individual events. Whether this (or other) probability attaches to an individual event does not depend on the availability of this assurance. Rather, it depends on the operation of the indifference principle and the extendibility presumption. These inferential rules apply to an individual event in the absence of information accounting for the difference between the cases in which the event occurs and the cases in which it does not occur. The reasoner will thus always make an epistemically unwarranted assumption that the unavailable information is not slanted in any direction.

The mathematical system may try to adopt a more demanding informational criterion: one that differentiates between probability estimates on the basis of their epistemic weights. *See* KEYNES, *supra* note 56, at 71–77. For contemporary analyses of Keynes's "weight" criterion, see COHEN, *supra* note 5, at 102–09; SCHUM, *supra* note 39, at 251–57; ALEX STEIN, FOUNDATIONS OF EVIDENCE LAW 80–91 (2005); L. Jonathan Cohen, *Twelve Questions About Keynes's Concept of Weight*, 37 BRIT. J. PHIL. SCI. 263 (1985). Charles Peirce also endorsed this criterion when he observed that "to express the proper state of our belief, not *one* number but *two* are requisite, the first depending on the inferred probability, the second on the amount of knowledge on which that probability is based." Charles Sanders Peirce, *The Probability of Induction*, 12 POPULAR SCI. MONTHLY 705 (1893), *reprinted in* 2 COLLECTED PAPERS OF CHARLES SANDERS PEIRCE 421 (Charles Hartshorne & Paul Weiss eds., 1932).

Under this criterion, the weight of a probability estimate will be determined by the comprehensiveness of what the reasoner does and does not know about her case. *See* KEYNES, *supra* note 56, at 71, 77. The decisional synergy between probability and weight will create a serious problem of incommensurability. Consider a reasoner who faces a high but not weighty probability, on the one hand, and a weighty but low probability, on the other hand. Which of the two probabilities is more dependable than the other? This question does not have a readily available answer. There is simply no metric by which to compare the two sets of probabilities. This problem may not be insurmountable, but why tolerate it in the first place? Why try hard to undo the damage caused by the mathematical system's epistemological outlaws, instead of barring them? Part III below responds to this question.

[70] For a prominent economist's conjecture as to what those causes might be, see Gilboa, *supra* note 12, at 6, who speculates that economists uniformly use mathematical probability because it is "theoretically clean: There is but one type of uncertainty and one way to model it." Economists' skepticism about mathematical probabilities can be traced back to Frank Knight, who famously called for a conceptual separation between "probability" as an estimate of calculable risks and "uncertainty"—a state of affairs to which no numerical estimate of probability can be assigned. *See* FRANK H. KNIGHT, RISK,

ber of serious distortions in the economic analysis of law, and I now turn to identifying those distortions.

The common baseline of all deterrence-driven doctrines is the general probability of law enforcement. When this general probability is too low, scholars of law and economics recommend an increase in the applicable penalty.[71] For example, when a legal rule is enforced only in one case out of ten, the penalty for its violators should be ten times greater than the penalty that the legal system would impose if it succeeded in enforcing the rule fully.[72] Courts and legislators often follow this recommendation.[73] All participants in this discourse ignore a simple fact of life: the law enforcement's general probability has virtually no effect on individual actors. Individual actors care about what affects them individually, not about what affects people in general. Correspondingly, those actors only care about their individual chances of receiving a penalty from the legal system. Basing their incentives upon mathematical probability is bound to create distortions. In

---

UNCERTAINTY AND PROFIT 19–20, 197–232 (1921). John Maynard Keynes attempted to connect the two probability systems—causal and mathematical—by introducing the concept of "weight": an imprecise measure of the wealth, or variety, of the evidence that forms the basis for numerical probability assessments. *See* KEYNES, *supra* note 56, at 71–77. Another renowned economist, G.L.S. Shackle, argued that any probability is conditional upon absence of "surprise"—a constant uncertainty condition that cannot be assessed in numerical terms. *See* G.L.S. SHACKLE, UNCERTAINTY IN ECONOMICS AND OTHER REFLECTIONS 1–16, 56–62 (1955); *see also* Tony Lawson, *Probability and Uncertainty in Economic Analysis*, 11 J. POST KEYNESIAN ECON. 38, 41–52 (1988) (explaining how three prominent economists—John Maynard Keynes, Frank Knight, and Robert Lucas—have criticized the empirically unfounded ascriptions of probability in economic theories and, in particular, the economists' conversion of uncertainty into mathematical probability); Marcello Basili & Carlo Zappia, *Ambiguity and Uncertainty in Ellsberg and Shackle*, 34 CAMBRIDGE J. ECON. 449 (2010) (relating Shackle's incorporation of "surprise" in probability assessments to Ellsberg's behavioral theory of ambiguity aversion).

71 For an overview of penalty increases, see sources cited *supra* note 4.

72 *See* sources cited *supra* note 4.

73 *See, e.g.*, Bank Holding Company Act, 12 U.S.C. §§ 1841, 1975 (2006) (imposing treble damages upon banks for engaging in prohibited nonbanking activities); Clayton Act, 15 U.S.C. § 15 (2006) (imposing treble damages for antitrust violations); Lanham Act, 15 U.S.C. § 1117(b) (2006) (imposing treble damages upon users of counterfeit trademarks); Consumer Credit Protection Act, 15 U.S.C. § 1693f(e) (2006) (imposing treble damages for violations of consumers' statutory rights); Racketeer Influenced and Corrupt Organizations Act, 18 U.S.C. § 1964(c) (2006) (imposing treble damages for RICO violations); Antiterrorism Act, 18 U.S.C. § 2333(a) (2006) (entitling victims of international terrorism to recover from a liable defendant "threefold the damages he or she sustains"); False Claims Act, 31 U.S.C. § 3729(a) (2006) (imposing treble damages on anyone who submits false claim for payment to federal agency); Patent Act, 35 U.S.C. § 284 (2006) (providing for treble damages for willful patent infringement); GA. CODE ANN. § 44-7-35(c) (2009) (allowing tenant to recover treble damages for landlord's improper withholding of rent deposit); MASS. GEN. LAWS ch. 231, § 85J (2010) ("Whoever, by deceit or fraud, sells personal property shall be liable in tort to a purchaser in treble the amount of damages sustained by him."); MISS. CODE ANN. § 89-7-25 (1999) (imposing double-rent liability on holdover tenants); N.J. STAT. ANN. § 2A:42-5 (West 2010) (same); Perez v. Z Frank Oldsmobile, Inc., 223 F.3d 617, 620–24 (7th Cir. 2000) (mentioning punitive and treble damages among mechanisms counteracting insufficient deterrence); *see also* MASS. GEN. LAWS ch. 221, § 51 (2010) ("An attorney at law who unreasonably neglects to pay over money collected by him for and in behalf of a client, when demanded by the client, shall forfeit to such client five times the lawful interest of the money from the time of the demand.").

section A below, I illustrate this distortionary effect by analyzing a recent decision of the Seventh Circuit,*United States v. Elliott*.[74]

Inattention to the incompatibility problem also has a profound effect on the economic analysis of torts. The extent of tort liability is determined by expected harm: the actual harm multiplied by the probability of its infliction.[75] The greater the expected harm, the greater the burden of precautions that prospective injurers have to take to avoid the harm.[76] Under the classic Learned Hand formula, an injurer assumes liability for the harm he inflicts on the victim when $B < PL$.[77] That is, the injurer is liable when his expenditure on precautions that could prevent the harm (the burden of precautions, denoted as B) is lower than the victim's loss (L) discounted by its probability (P).[78] The harm's probability consequently becomes a key factor in liability analysis and decisions. Scholars of law and economics uniformly endorse the mathematical understanding of the harm's probability.[79] They associate this probability with instantial multiplicity—a criterion that focuses on the general incidence of harm-causing accidents across cases.[80] This approach has found its way into several court decisions that have applied the Learned Hand formula.[81]

This approach leads a prospective injurer astray by prompting him to merge his case-specific information with the general accident statistics, and to produce the rate for the type of accident that he should either prevent (if $B < PL$) or let happen (if $B \geq PL$). As I have already explained, this procedure trims away the individual characteristics of the injurer's case. The statistical figure it delivers in the end will give the injurer the average rate of accidents in cases similar to his, but will hardly say anything informative about the case at hand. This individual case may actually be on the high end of the statistical spectrum. Alternatively, it may actually be on the low end or close to the middle. The injurer therefore will do better by attaching crucial significance to his case-specific information about the relevant causes and effects. More importantly, society would be better off if prospective injurers were to base their actions on such information. Reliance on naked

---

[74] 467 F.3d 688 (7th Cir. 2006).

[75] *See* POSNER, *supra* note 3, at 167–69.

[76] *Id.*

[77] *Id.*

[78] *Id.*

[79] *Id.*; *see also* CALABRESI, *supra* note 10, at 255–59 (favoring a statistical approach over case-by-case assessments of a harm's probability in the context of accident costs' minimization); SHAVELL, *supra* note 10, at 177–93 (providing a basic overview of liability and deterrence theory based in mathematics).

[80] *See, e.g.*, POSNER, *supra* note 3, at 169–71 (explaining negligence standards by reference to general probability and statistical averages); SHAVELL, *supra* note 10, at 177 (assuming that "accidents and consequent liability arise probabilistically").

[81] *See, e.g.*, POSNER, *supra* note 3, at 169–70 (citing court decisions that relied on the Hand formula and similar reasoning in making negligence determinations).

statistics can only minimize actuarial harm that exists on paper. Case-specific causal analysis is likely to prevent actual harm.

The "level of activity" theory[82] illustrates a different aspect of the incompatibility problem and its neglect by tort scholars. This theory uses a purely statistical association between the level of a risky activity and the resulting harm as a basis for far-reaching policy recommendations. The case-specific causal analysis I advocate in this Article rejects this association.

A. Penalty Multipliers

A recent application of the penalty-multiplier doctrine took place in a case involving a convicted criminal who failed to report to prison to begin a five-year sentence. The criminal fled to Arizona, where he lived free under a borrowed identity for fifteen years.[83] He was then apprehended by the FBI and brought to trial. His guilt was not in dispute, but his sentence presented a number of issues that required the Seventh Circuit's intervention.[84]

Writing for the Circuit, Judge Easterbrook decided that the criminal's punishment for absconding should offset his expected gain from that crime.[85] He estimated that the criminal converted his original sentence to an imprisonment postponed by fifteen years with "a substantial [50%] chance that it would never start at all."[86] Judge Easterbrook also determined that the flight allowed the criminal to expedite the enjoyment of freedom which he could lawfully enjoy only after serving five years in jail. He ruled in that connection that "[t]ime served in future years must be discounted to present value"[87] and that "a modest discount [of] 5% per annum"wasappropriate.[88] Based on those baseline assessments, Judge Easterbrook calculated that the criminal "evaded 75% of the deterrent value of his five-year sentence"[89] and that this evasion—forty-five months of jail time—wasthe criminal's ill-gotten gain from the flight.[90] Judge Easterbrook remanded the case to the district court with an instruction to factor this gain into the criminal's punishment for absconding.[91]

---

[82] *See* SHAVELL, *supra* note 10, at 193–99 (articulating the "level of activity" theory). For the classic account, see Steven Shavell, *Strict Liability Versus Negligence*, 9 J. LEGAL STUD. 1 (1980).

[83] United States v. Elliott, 467 F.3d 688, 689 (7th Cir. 2006).

[84] *Id*. at 691 ("Because the district judge miscalculated the Guideline range, which he used as a starting point, the error may have affected Elliott's sentence, and we must remand. This does not imply, however, that a sentence of 21 months is unreasonably high; to the contrary, it strikes us as unreasonably low, and *United States v. Booker*, 543 U.S. 220 (2005), gives the district court ample authority to impose an appropriate sentence on remand." (citation omitted)).

[85] *Id*. at 691–92.

[86] *Id*. at 692.

[87] *Id*.

[88] *Id*.

[89] *Id*.

[90] *Id*.

[91] *Id*. at 693.

In the paragraphs ahead, I question Judge Easterbrook's assessment of the criminal's probability of avoiding the sentence. The 50% figure represents Judge Easterbrook's estimation of the percentage of criminals who successfully run away from the law.[92] This estimation is rough, but its roughness is not what I focus upon here. What I focus upon is the nexus between this general statistic and the individual defendant, Mr. Elliott. Did Mr. Elliott generate for himself a 50% chance of successfully evading his prison sentence? For Judge Easterbrook, this question was manifestly obvious. From the mathematical probability perspective that he adopted, the average criminal's probability of successful absconding attaches to all runaways, including Mr. Elliott. Hence, Mr. Elliott did create for himself a 50% chance of staying free instead of going to jail.

The attribution of a 50% chance to Mr. Elliott relies on the principle of indifference. Under this principle, absent special reasons for distinguishing between different runaways, all runaways should be treated as equals. Assuming, again, that one runaway out of two is eventually caught, an average runaway misappropriates 50% of the freedom with which he ought to pay for his prior crime. Each runaway's punishment for absconding therefore should be enhanced by half of his sentence for the prior crime. Together with Judge Easterbrook's present-value adjustment, this enhancement would deter criminals from running away from the law.

This approach follows the penalty-multiplier rule, designed by Professor Gary Becker.[93] This rule holds that the penalty that a criminal should receive must equal the penalty that he would have received in a world in which law enforcement is perfect (P) divided by the probability that the legal system actually delivers that penalty.[94] This probability equals the fraction of cases ($1/q$) in which the legal system delivers the penalty to criminals. Under the penalty-multiplier rule, the fraction's denominator (q) functions as a multiplier that aligns the criminal's expected penalty ($1/q \times P$) with the ideal penalty (P).[95] Introduction of this multiplier will induce the criminal to act in the same way in which he would have acted if his punishment for the crime were certain. This means that a legal system experiencing drawbacks in law enforcement need not expend money and resources in order to fix those drawbacks. All it needs to do is to up the penalty to the appropriate level—a measure it can implement with a strike of a pen.

---

[92] Judge Easterbrook estimated that Mr. Elliott had "evaded 75% of the deterrent value of his five-year sentence" in the following way: "As a deterrent, a 50% chance of serving five years starting 15 years from now must have less than 25% the punch of five years, with certainty, starting right now. This represents only a modest discount (about 5% per annum); many people discount the future even more steeply." *Id*. at 692.

[93] *See* Becker, *supra* note 4, at 180. The basic idea can be traced back to JEREMY BENTHAM, AN INTRODUCTION TO THE PRINCIPLES OF MORALS AND LEGISLATION 170 & n.1 (Hafner Publ'g Co. 1948) (1823).

[94] *See* Polinsky & Shavell, *Economic Analysis*, *supra* note 4, at 889–90.

[95] *Id.*

According to Judge Richard Posner's succinct formulation of the same idea:

> If the costs of collecting fines are assumed to be zero regardless of the size of the fine, the most efficient combination is a probability arbitrarily close to zero and a fine arbitrarily close to infinity. . . . [E]very increase in the size of the fine is costless, while every corresponding decrease in the probability of apprehension and conviction, designed to offset the increase in the fine and so maintain a constant expected punishment cost, reduces the costs of enforcement—to the vanishing point if the probability of apprehension and conviction is reduced arbitrarily close to zero.[96]

This measure has a serious handicap: its underlying assumption that the general probability of law enforcement defines an individual criminal's expectation of penalty is false.[97] A criminal's expectation of penalty is defined by her individual probability of being caught and punished. This individual probability crucially depends on the law enforcers' factual proximity to the criminal. This proximity is not determined by the general rate of law enforcement. Rather, it is determined by a combination of case-specific factors *causatively relevant* to the criminal's apprehension and punishment as an empirical matter. The number and variety of those factors are therefore the only information that the criminal would rationally care about.

Assume, hypothetically, that a legal system decides to adopt Judge Posner's model (after finding the way to eliminate the fine collection problem). It prescribes an astronomic fine for the crime in question and reduces the number of law enforcers on its payroll to ten people, who randomly check on suspects. Holmes's "Bad Man"[98] contemplates the commission of a crime after finding no evidence confirming the individual scenario in which one of those ten law enforcers apprehends him. Bad Man only finds out that the law enforcers exact the fine from one offender out of 500,000. How will he respond to this statistical information?

---

[96] *See* POSNER, *supra* note 3, at 221.

[97] In a recently published book, Judge Posner embraces skepticism about abstract mathematical probabilities. He convincingly argues that economists oftentimes substitute nonquantifiable uncertainties by empirically meaningless probabilistic figures and that this epistemic error, along with the regulators' failure to mitigate the uncertainty problem, is partly responsible for the current financial crisis. *See* RICHARD A. POSNER, THE CRISIS OF CAPITALIST DEMOCRACY 288–304 (2010). Judge Posner's discussion draws on the works of Frank Knight and John Maynard Keynes, referenced (together with other relevant sources) in notes 69–70 above. Judge Posner impliedly rejects the "indifference principle"—a pillar of the mathematical probability system that substitutes unknown distributions with equal distributions. *See supra* notes 54–59 and accompanying text. Whether Judge Posner is ready to accept the full Keynesian package of "probability" and "weight," *see supra* notes 69–70, is unclear.

[98] *Cf.* O.W. Holmes, *The Path of the Law*, 10 HARV. L. REV. 457, 459 (1897) (famously defining the "Bad Man" as a person "who cares nothing for an ethical rule which is believed and practised by his neighbors [but] is likely nevertheless to care a good deal to avoid being made to pay money, and will want to keep out of jail if he can").

Bad Man will certainly mind this information, especially if he asks for a statistician's advice. However, he will then be equally mindful of another statistical fact: his probability of being struck by lightning reportedly equals 1/400,000.[99] Bad Man's fear of conviction and punishment consequently would be offset by the prospect of being killed by a lightning before being caught by the police. For a devout statistician, this setoff may be unproblematic. If so, the statistician needs to be reminded of all other low-probability fatalities. Taking those fatalities' probabilities into account will change Bad Man's situation quite dramatically. Under the disjunction rule, the aggregated effect of those fatalities equals the weighted sum of their probabilities. This sum will steeply reduce Bad Man's chances of staying alive when the law-enforcers knock on his door. Under those circumstances, only a credible threat of a painful afterlife punishment will induce Bad Man to stay away from crime.

In reality, of course, Bad Man will not pay much attention to his actuarial death. His passing away has no individual causative confirmations (besides the fact of mortality that attaches to all humans). By the same token,Bad Man will not pay much attention to his actuarial prospect of paying the high fine. To make him actually fear this prospect, the legal system must set up mechanisms for apprehending criminals in a nonaccidental way and show that those mechanisms actually work. This measure might bring the law enforcement prospect close enough to Bad Man's doorstep. Threats on paper, however, will not suffice.

Consider now the specifics of the *Elliott* case.[100] These specifics include Mr. Elliott's experience and sophistication as an attorney and a partner in a prestigious law firm, as well as his ability to recruit a reliable collaborator—a cousin—who allowed him to use his name and other personal information in order to start a new life as a fugitive in Arizona.[101] Mr. Elliott's ex ante probability of being brought to justice was therefore far below 50%. He secured his escape and change of identity and disappeared into Arizona. There, his fugitive life went virtually undisturbed by an individuated causative prospect of apprehension. Thus, Mr. Elliott's individuated-causative probability of not paying for his crime was very high.

B. Actuarial Harms

The economic interpretation of the Learned Hand formula requires prospective injurers to perform a fairly complicated probabilistic calculation.[102] An injurer first needs to ascertain the harm's prior probability: the probability that the victim will sustain harm if the injurer takes no precau-

[99] *See* NAT'L OCEANIC & ATMOSPHERIC ADMIN., NAT'L WEATHER SERV. LIGHTNING SAFETY MEDICAL INFORMATION, http://www.lightningsafety.noaa.gov/medical.htm (last visited Nov. 23, 2010).

[100] United States v. Elliott, 467 F.3d 688, 689 (7th Cir. 2006).

[101] *Id.*

[102] *See, e.g.*, ROBERT COOTER & THOMAS ULEN, LAW & ECONOMICS 336–38 (5th ed. 2008) (detailing the calculus required for application of the Hand formula).

tions against that harm. The injurer should determine this probability by relying on general experience. The injurer subsequently needs to multiply this probability by the estimated total amount of the harm. The resulting sum will determine his maximal expenditure on precautions against the harm. The injurer then needs to compile a list of available precautions, hypothesize that he takes those precautions both individually and conjunctively, and calculate the harm's probability for each scenario. As previously noted, this calculation will rely upon general experience with similar accidents and precautions. Finally, the injurer must calculate the difference between the harm's prior probability, on the one hand, and the harm's probability under each precautionary scenario, on the other hand. This difference, multiplied by the harm, will determine the benefit of each precautionary measure. The difference between each of those benefits and the cost of the precautionary measure that produces the benefit will determine the measure's utility. Among the available precautionary measures, the injurer should choose the one that produces the greatest utility. By acting in this way, he will minimize the harm and the harm-preventing expenditures as a total sum.[103]

Unfortunately, this economic prediction is true only in the actuarial sense. For reasons described in Part I, the injurer's calculations will only give him mathematical averages. Those averages replicate the underlying empirical facts in the same way in which 4 replicates 3 and 5. Those averages will not necessarily produce a bad decision. Yet they virtually never correspond to the individual causes and effects that determine whether the injurer's precautions will actually prevent the harm. These causes and effects are parts of individual cases that collectively determine the mathematical average. The reverse, however, is not true: a mathematical average never determines what will happen in the individual cases of which it is composed. An injurer therefore always needs to make a nonstatistical evaluation of causes and effects that are present in his specific case. This evaluation will give him a better sense of his likelihood of causing harm to another person.

To illustrate, consider a person driving her car behind another vehicle and trying to calculate the probability of a rear-end collision with that vehicle. The driver, of course, can use a rule of thumb instead of probability: for example, she can rely on the rule that instructs drivers to maintain a four-second following distance between cars. The driver, however, wants to use mathematical probability, and I assume arguendo that she can calculate this probability correctly in a blink of an eye. The driver gathers the relevant information from the road and properly combines it with the general accident statistics. Her bottom-line figure, 0.5, informs her that she has a

---

[103] *Id. But see* Ronald J. Allen & Ross M. Rosenberg, *Legal Phenomena, Knowledge, and Theory: A Cautionary Tale of Hedgehogs and Foxes*, 77 CHI.-KENT L. REV. 683 (2002) (arguing that courts do not require real-world people to make such calculations and providing examples).

50% chance of colliding with the vehicle she follows if both vehicles keep on driving at the same speed. What exactly does this statistic mean to her?

It means that, on average, when all cases in the statistical sample are deemed equal, one case out of two involves a rear-end collision. Also, since the driver's case exhibits no features separating it from all other cases in the sample, this case too has a 50% chance of collision. The sampled cases, however, are actually not identical. Far from it: the driver knows that half of those cases do not involve collisions. If so, how should the driver distinguish between the two categories of cases? The mathematical probability theory advises the driver not to distinguish between those categories because she has no information for making that distinction. Specifically, the theory advises the driver to apply the principle of indifference and deem all the cases in her sample equally likely to involve a car collision.

But why make this counterfactual assumption? Why not bring into play more case-specific factors that have, or may have, causal significance? For example, why not allow the driver to consider her individual driving skills, as well as the skills exhibited by the driver of the vehicle she follows? Indeed, why not advise the driver to take into account the width of the road's shoulder onto which she could swerve in the event of emergency? More fundamentally, why not instruct the driver that, instead of relying on mathematical probability, she should consider which of the two scenarios that affect her individually—her car's involvement and noninvolvement in a rear-end collision—has the strongest causative confirmation? In short, why not advise the driver to switch from indifference to difference and base her decision upon evidential variety, instead of instantial multiplicity? The accident's mathematical probability allows the driver to make an intelligent gamble, but this is not good enough. By contrast, case-specific evaluation of the relevant causal indicators enables the driver to make an informed assessment of her individual risk and to adequately respond to that risk.

C. The "Level of Activity" Theory

Scholars of law and economics complain that the conventional negligence doctrine is not sufficiently probabilistic. Specifically, they argue that the doctrine focuses exclusively on the injurer's level of care while neglecting what they perceive to be an important dimension of tortious risk: the level of the injurer's risky activity. The negligence doctrine, so goes the argument, incentivizes injurers to exercise adequate care, while allowing them to repeat risky activities as much as they please so long as the requisite precautions are taken. For example, a person who drives her car carefully enough can drive it as often as she wants. The law thus allows the person to drive her car unnecessarily. The person's driving consequently may increase the probability of a car accident without generating offsetting benefits for society.

In numerical terms, when a reasonable precaution guaranteeing an exemption from liability costs the injurer $6 per unit of her risky activity, and her gain from that activity is $7 per unit, the injurer will engage in the activity even when each additional unit increases the victim's expected damage to $10. Doing so willyield the injurer a $1 profit (the $7 gain minus the $6 expenditure on the legally required precaution). At the same time, society's welfare—about which the injurer does not care—will decrease by $9 (the difference between the victim's $10 damage and the $1 profit that the injurer must give up to avoid the damage).

Law and economics scholars argue that the law should rectify this misalignment between the injurer's interests and society's good. They make two recommendations that aim at achieving this result. First, they call for the replacement of the negligence doctrine by a strict liability regime.[104] Second, and more ambitiously, they urge the government to regulate the levels of risky activities.[105]

The "level of activity" theory relies on the disjunction rule. Under this rule, multiple possibilities of an accident steadily increase the probability of the accident's occurrence. Based on this actuarial truism, the level of activity theory argues that since driving a car always involves the possibility of an accident, a person who repeatedly drives his car increases the probability of accidents.[106] Consequently, according to this theory, the law should regulate that person's driving even when he drives with adequate care on every individual occasion. Otherwise, the person willintensify his driving activity and increase the probability of an accident for no good reason. For instance, the person might hit the road just to show off his new car or to buy gourmet food for his pet iguana.[107]

This theory ignores a simple truth about causation: the amount of a risky activity can never cause damage. Damage is never inflicted by multiple repetitions of the same activity that causes harm once in a while. Rather, it is inflicted by one specific activity that inflicts harm on a specific victim, incrementally or in one shot. This damaging activity, indeed, may well be the very first in a series of actions taken by the injurer. More fundamentally, a person's repeated, but invariably cautious, driving on a highway can *cause* a car accident no more than the tree or a utility pole into which careless drivers occasionally collide. The torts system therefore needs to deter only careless, rather than frequent, car drivers, as it actually does.

Under the negligence doctrine, the injurer willbe liable for the victim's damage if the court finds that he could have avoided the damage by taking

---

[104] *See* SHAVELL, *supra* note 10, at 196.

[105] *See* David Gilo & Ehud Guttel, *Negligence and Insufficient Activity: The Missing Paradigm in Torts*, 108 MICH. L. REV. 277, 288–89, 302–03 (2009) (analyzing regulatory ways of reducing the level of risky activities).

[106] *See* SHAVELL, *supra* note 10, at 193–95.

[107] *See* POSNER, *supra* note 3, at 178.

precautions that are not disproportionately costly given the damage's magnitude and probability. Under strict liability, in contrast, the injurer assumes liability when the court determines that he was the cheapest avoider of the damage. If the court finds that the victim was best positioned to avoid his own damage, the victim willonly recover partial compensation or no compensation at all.[108] Under both regimes, the court's inquiry will focus upon individual causation rather than statistical correlation. Courts consequently will impose no liability on a person who takes his luxury car to a highway every day and carefully drives it for five hours in order to spur envy from other drivers. If that person gets involved in an accident caused by another driver's negligence, the negligent driver willnot be allowed to blame any part of the accident and the resulting damage on the other driver's unnecessarily intensive—and, arguably, ill-motivated—presence on the road. The law that dictates this result is both fair and efficient.

For these reasons, the "level of activity" theory has always been—and will likely remain—just a theory.[109] Courts have never used the high level of an injurer's activity as a reason for holding him liable in torts.[110] Nor have they treated a low level of an injurer's activity as an exonerating circumstance.[111] The level of an injurer's risky activity can only be taken into account in estimating the cost of precautions against the victim's damage. When the injurer performs the same activity repeatedly, it becomes cheaper for him to set up a durable precaution against damage. The cost of this precaution will then be spread across many activities, as opposed to just one.[112] Courts account for this economy of scale under both negligence and strict liability regimes.[113] The negligence doctrine therefore has no flaws that the level of activity theory can fix.

As I already mentioned, the level of activity theory also forms the foundation for an ambitious regulatory proposal. This proposal holds that the government should regulate the frequency of activities that may end up in an accident. Would it be a good idea for the government to adopt this proposal and start regulating excessive driving in addition to unsafe driv-

---

[108] *Id.* at 172–75.

[109] *See* Kenneth S. Abraham, Response, *Insufficient Analysis of Insufficient Activity*, 108 MICH. L. REV. FIRST IMPRESSIONS 24, 25 (2009), http://www.michiganlawreview.org/assets/fi/108/abraham.pdf (attesting that the traditional negligence and strict liability rules adequately solve the problems that the level of activity theory struggles with).

[110] *See id.* at 27.

[111] *See* Richard A. Epstein, Response, *Activity Levels Under the Hand Formula: A Comment on Gilo and Guttel*, 108 MICH. L. REV. FIRST IMPRESSIONS 37, 40 (2009), http://www.michiganlawreview.org/assets/fi/108/epstein.pdf.

[112] *See* Mark F. Grady, *Why are People Negligent? Technology, Nondurable Precautions, and the Medical Malpractice Explosion*, 82 NW. U. L. REV. 293, 302 (1988).

[113] The opposite, however, is not true. An injurer cannot be held liable for failing to take a disproportionately expensive precaution against the victim's damage on the theory that he could have reduced the precaution's marginal cost by intensifying his risky activity. *See* Abraham, *supra* note 109, at 26–27.

ing? For example, should a regulator introduce mandatory carpools, road tolls, and surtaxes on gasoline as an incentive for people to cut back on driving their vehicles?

I posit that it would be a bad idea. Consider again the proposition that excessive driving raises the probability of accidents even when it is safe. This proposition is correct in the sense that the incidence of accidents as a total number increases with the number of interactions on the road. At the same time, however, a *safe* driver decreases the incidence of accidents per each unit of the driving activity. Given the presence of unsafe driving, every driving of a vehicle by a safe driver will have this statistical effect.

This effect is not a good reason for encouraging safe drivers to congest roadways in order to show off their cars or satisfy the culinary cravings of their iguanas. Yet, underscoring it brings about a methodological benefit: the effect's presence positions causation at the center of policymakers' attention. Reducing the volume of safe driving will not necessarily prevent accidents because accidents' occurrence crucially depends on what *unsafe* drivers do. Safe driving is a mere background condition for accidents caused by unsafe drivers.[114] Arguably, the total number of accidents can be brought down by the dilution of the general accident opportunity. This actuarial prediction, however, is not causatively robust because the opportunity to cause accidents is not equally distributed across drivers. Unsafe drivers seize upon that opportunity, while safe drivers avoid it.

Heterogeneity of individuals' driving capabilities constitutes a compelling reason for *not* regulating the level of the driving activity.In the best possible scenario, such regulation will restrain all drivers, those who drive safely and those who do not. The equal imposition of the regulatory constraint will create an anomalous cross-subsidy: safe drivers will have to sacrifice part of their driving-related benefits in order to downsize the unsafe drivers' accident opportunity. The prevalent torts doctrine precludes this cross-subsidy by tying the "driving tax"—i.e., the duty to pay for the harm negligently caused—to the safety of each individual driver. To strengthen this tie, the doctrine disconnects itself from naked statistical correlations and relies upon individuated causative indicators that vary from one case to another. This doctrine is not broken and need not be fixed.

III. CAUSATIVE PROBABILITY

My preceding discussion has outlined the defining characteristics of the causative system of probability. In this Part of the Article, I specify the system's details and explain its fundamental disagreements with the mathematical system. This discussion proceeds in two sections. Section A explains the system's distinct logic. Section B sets forth and elucidates the

---

[114] Safe driving may raise the incidence of unavoidable car accidents. Those accidents, however, are both rare and too costly to avoid. Their probability therefore cannot be a good reason for inducing safe drivers to stay off the road.

system's epistemic principles. Both sections demonstrate that the causative system of probability maximizes a person's epistemic grasp of individual events.

The system's logical makeup is best described by what I call the "difference principle." I have chosen this name for a number of reasons. The difference principle conceptualizes the operational gap between the causative system of probability and the mathematical system that guides itself by the principle of *in*difference. The causative system associates probability with the extent to which the reasoner's information confirms and disconfirms the occurrence of the relevant event. The difference between those conflicting evidentiary confirmations determines the event's probability. The mathematical system, in contrast, postulates—artificially—that the reasoner's information is complete and then identifies the event's probability with the instantial multiplicity that is present in that information. This method of reasoning assumes that information not available to the reasoner is not slanted in any direction and therefore does not make a difference. The causative system, in contrast, evaluates the difference that the unavailable information would have made if it were available.

Section A demonstrates that the difference principle resonates with Mill's methods of "difference" and "agreement" that allow reasoners to determine causative probabilities of individual events.[115] This principle also reflects Bacon's "elimination method"[116] and his foundational insight that any extraction of facts from a multiplicity of events can be falsified by a single occurrence of a different event.[117] The epistemics of the causative probability system are driven by its evidential-variety criterion. As I have already explained, this criterion requires the reasoner to analyze her information by considering the relevant causal indicators: those that confirm the occurrence of the event under consideration and those that point in the opposite direction. The reasoner must carry out a comparison between those indicators based on their number and scope.

In section B below, I advance the understanding of this criterion by applying it to a number of cases by which I previously illustrated the failings of mathematical probability. I show that this criterion's application always produces a decision that best suits the reasoner's individual case.

A. The Difference Principle

The difference principle originates from Bacon's famous observation about the epistemic limit of instantial multiplicity. Bacon wrote that no number of favorable instances can establish the epistemic validity of a ge-

---

115 *See* MILL, *supra* note 20, at 278–91.
116 *See* BACON, *supra* note 21, at Book I, Point 46, at 221.
117 *Id.*

neralization; yet, a single instance unfavorable to a generalization can invalidate it.[118]

Take a rural road that is virtually never patrolled by the police, and consider the probability of a speeding driver's apprehension on that road. The mathematical probability of that scenario will obviously be next to zero. But what does this probability mean to the same driver tomorrow? Not much, because tomorrow is literally another day. Tomorrow, the police may actually patrol the road. The driver therefore willhave to look out for police presence on the road before she decides to speed. If she encounters a police patrol, her individual case willinvalidate the no-enforcement generalization. On the other hand, if the driver encounters no police presence and decides to speed, her case willcoincide with the no-enforcement generalization but willnot confirm it. This generalization willreceive no confirmation because the driver's decision to speed willnot rely on the number of past occasions on which the road was free of police presence. Rather, it willrely on the driver's event-specific elimination of the apprehension risk. The driver willreason in the following way: "There are no police cars on this road today. Therefore, the police will not apprehend me." This form of reasoning is what causative probability is about.

Bacon's mistrust of instantial multiplicities led him to develop the "elimination method."[119] This method systematically prefers proof by elimination to evidential confirmation. According to Bacon, information associated with a particular event—no matter how extensive it is—cannot establish that this event will occur. To establish that an event is probable, the reasoner needs to have information that eliminates rival possibilities. The scope of the eliminating information determines the event's probability.[120] Thus, the event's probability increases as more and more alternative scenarios are eliminated.[121] Complete elimination of the rival possibilities will establish the event's occurrence with practical certainty.

The driver in my example can benefit from this procedure as well. After finding no police patrol on the road, she should try to eliminate other scenarios contradicting her no-apprehension hypothesis. For example, she should consider other drivers' behavior on the road. If those drivers are

---

[118] This point is summarized in Bacon's celebrated phrase, "Major est vis instantiae negativae." *Id.* In the same paragraph that coined this phrase, Bacon sharply criticizes the widespread preference of affirmations over negations, describing it as an "intellectual error." *See also id.* at n.67 (commentary of editor Thomas Fowler) ("A single negative instance, if it admit of no explanation, is sufficient to upset a theory, or, at the least, it ought to cause us to suspend our judgment, till we are able either to explain the exception, or to modify the theory in accordance with it, or else to accumulate such an amount of negative evidence as to justify us in rejecting the theory altogether. The negative instance, even where it does not upset a theory, is often peculiarly valuable, in calling attention to a counteracting cause."); COHEN, *supra* note 5, at 4–13 (discussing Bacon's tradition in the philosophy of induction); KNEALE, *supra* note 36, at 48–53 (analyzing Bacon's method of induction by elimination).

[119] *See* COHEN, *supra* note 5, at 145–56.

[120] *Id.*

[121] *Id.*

speeding as well, it willbe safe for her to assume that they, too, do not see police vehicles in their vicinity. The driver may also rely on the radar detector device in her car. The device's silence would indicate that there are no radar-equipped cars on the road.

To complete my outline of Bacon's method, I now modify the example. Assume that, on account of scarce resources, the police decided not to monitor drivers on the road in question, and our driver learns about that decision. Based on this information, she decides to speed. From Bacon's point of view, this set of facts fundamentally differs from the previous example. Under the present set of facts, empirically identifiable causal forces (police chiefs) have removed police patrol cars from the road. Absence of police monitoring consequently becomes an empirically established fact. This fact negates, if not altogether eliminates, the enforcement possibility for all drivers, including ours. This negation validates the no-enforcement generalization.

Bacon's method has the virtue of identifying nonaccidental connections between causes and effects.[122] In my first example, this connection is formed between the police's absence and the driver's decision to speed. My second example illustrates a different causal connection: the connection between the police decision not to monitor drivers on the road and the driver's decision to speed. Modern philosophers brand those connections as law-bound (or law-like) regularities in order to distinguish them from coincidences.[123] Among these philosophers, Bacon is widely regarded as a founder of the modern scientific method.[124] Building on Bacon's approach, John Stuart Mill formulated a set of canons for determining causative probability (also identified as inductive or Baconian).[125]

Of those canons, the methods of "difference" and "agreement"[126] are most important.[127] These methods are best understood with the help of examples. Begin with the method of agreement. Consider four speeding drivers on the same rural road who slow their cars down more or less simultaneously. What could be the cause of this collective slowdown? The experiences of each driver capable of explaining his or her decision to slow down are listed in the table below:

---

122 *See id.* at 5–6.

123 *See supra* note 62 and sources cited therein.

124 *See, e.g.*, FRIEDEL WEINERT, COPERNICUS, DARWIN AND FREUD: REVOLUTIONS IN THE HISTORY AND PHILOSOPHY OF SCIENCE 148–51 (2009); LISA JARDINE, FRANCIS BACON: DISCOVERY AND THE ART OF DISCOURSE 3 (1974) (describing Bacon's NOVUM ORGANUM as "a forerunner of modern scientific method").

125 *See* COHEN, *supra* note 5, at 145 (describing causative probability as Baconian); COHEN, *supra* note 23, at 121–23 (describing causative probability as inductive).

126 *See* MILL, *supra* note 20, at 278–91.

127 *See* COHEN, *supra* note 23, at 144–51 (underscoring the centrality of those methods for causal inquiries).

driver \ experience	saw police patrol	noticed speed limit	felt tired	slowed down	feared collision with another car
A	YES	YES	YES	YES	NO
B	YES	YES	NO	YES	YES
C	YES	NO	NO	YES	NO
D	YES	NO	YES	YES	YES

Mill's method of agreement holds that "If two or more instances of the phenomenon under investigation have only one circumstance in common, the circumstance in which alone all the instances agree, is the cause (or effect) of the given phenomenon."[128] In the present example, the only experience that is common to all drivers was seeing a police vehicle patrolling the road. This factor, therefore, is the most probable cause of the drivers' slowdown (the investigated phenomenon). Note that the method of agreement incorporates Bacon's elimination procedure.[129] The reasoner lists the phenomenon's causal explanations and removes from her inquiry the epistemically inferior ones: those that explain some instances of the phenomenon but not others.[130] The remaining explanation—one that covers all instances of the phenomenon—consequently acquires epistemic superiority, and the reasoner is advised to treat it as the phenomenon's most probable cause.

To illustrate the method of difference, assume that Driver *A* did not slow her car down and that the drivers' experiences at that point in time included the following:

driver \ experience	saw police patrol	noticed speed limit	felt tired	feared collision with another car	slowed down
A	NO	YES	YES	YES	NO
B	YES	YES	YES	YES	YES
C	YES	YES	YES	YES	YES
D	YES	YES	YES	YES	YES

[128] MILL, *supra* note 20, at 280 (italics omitted).
[129] Mill acknowledged it explicitly. *See id.* at 281.
[130] *Id.* at 282–83.

The difference method applies to factual setups, one of which exhibits a certain effect (here, the driver's failure to slow down) while others do not. If all but one of those setups' circumstances are identical to each other, and the exceptional circumstance belongs to the setup in which the effect occurred, then this circumstance—"the difference"—is the most probable cause of the effect. In the present scenario, the only difference between Driver *A* and all other drivers is Driver *A*'s failure to see the police patrolling the road. This failure is the most probable cause of that driver's decision not to slow down.

The two methods can be applied both individually and in combination with one another.[131] They can also be adjusted in order to work with concurrent variations that are often discernible from a range of cases that exhibit a certain common effect. When this effect has a feature or property that varies concurrently with some factor that is present in every case, this factor is the most probable cause of the observed effect.

Mill called this adjustment "the method of concomitant variation."[132] To illustrate this method, hypothesize that the four speeding drivers in my present example are stopped by the police. The police officer then suspends Driver *A*'s license forthwith and issues fines to Drivers *B*,*C*, and *D* in the respective amounts of $100, $200, and $300. The observed effect here is the officer's reaction to the drivers' behavior on the road that varies in its severity from one driver to another. The effect's cause—the drivers' speeding—varies concomitantly with that effect.

The system of Bacon and Mill requires the reasoner to identify the information causally relevant to her hypothesis and base her decision on that information alone. The information's breakdown into "causally relevant" and "causally irrelevant" factors must follow Mill's methods. Subsequently, the reasoner ought to compare the information that confirms her hypothesis (causal positives) with the information that rejects it (causal negatives). This comparison must focus on the number and variety of causal positives, on the one hand, and causal negatives, on the other hand. The reasoner ought to carry out an epistemic assessment of those causal indicators in order to determine which of them provides the most extensive coverage for its underlying scenario. This criterion will determine the winner of the epistemic contest and the probability of the competing scenarios.

The causative system of probability differs from the mathematical system in virtually every material respect. On the most fundamental level, the causative system rejects the indifference principle, upon which the entire mathematical system rests. The causative system associates probability with the scope of the informational coverage for the relevant scenario. This criterion focuses on the size of the gap between the existing informational

[131] *Id.* at 284–85.
[132] *Id.* at 287–91.

coverage and the complete information. For inquiries guided by this criterion, the key question is how significant this gap is. The reasoner asks herself a factual question: would the missing information make a significant change in my decision, if it were to become available? The mathematical system, on the other hand, tells the reasoner that, instead of asking this difficult and possibly intractable question, she ought to simplify her task by assuming—counterfactually—that the unavailable information makes no difference. This convenient counterfactual assumption also allows the reasoner not to worry much about the size of the gap between her information and full information. To dispel the reasoner's worries about this gap, the system advises her to assume—again, counterfactually—that the unknown information is not slanted in either direction and that the probabilities that this information is associated with cancel each other out. These assumptions abstract the reasoner away from her actual case, in which the missing information *is* slanted because the event in question either will occur or will not. The mathematical system thus advises the reasoner to base her decision on the so-called average case—a theoretical construct that does not exist in the empirical world.

This advice sharply separates the two systems. As I already explained, the causative system of probability tells the reasoner to focus on her individual case and helps her identify the direction in which the missing information might go. This system rejects the mathematical system's indifference toward—and the consequent trivialization of—informational deficiencies. Those deficiencies do make a difference.

The causative system also rejects the uncertainty principle—a questionable epistemic device by which the mathematical system ascribes probabilities to completely unevidenced scenarios. The causative system gives no probabilistic credit to scenarios that have no evidential support.

On similar grounds, the causative system refuses to treat instantial multiplicities (and statistical generalizations deriving therefrom) as per se extendible. The mere fact that most events that appear indistinguishable from each other exhibit a particular characteristic is not a good reason for expecting this characteristic to be present in a new similarly looking event. Only a proven causal explanation or theory can establish a feature's extendibility across different events. Naked statistics will not do. Supporters of statistical inferences often say that, in the presence of uncertainty, all inferences are statistical.[133] However, this saying is profoundly mistaken. It relies on the fact that every inference requires a generalization and then goes on to

[133] *See, e.g.*, United States v. Veysey, 334 F.3d 600, 605–06 (7th Cir. 2003) ("'All evidence is probabilistic—statistical evidence merely explicitly so' . . . . Statistical evidence is merely probabilistic evidence coded in numbers rather than words.") (quoting Riordan v. Kempiners, 831 F.2d 690, 698 (7th Cir. 1987)).

suggest that, because all generalizations are statistical in nature, then every inference is statistical as well.[134]

Fortunately for all of us, not every generalization is statistical.[135] Some generalizations are statistical, while others are causal.[136] Statistical generalizations are extractable from each and every instantial multiplicity, but causal generalizations have a more solid epistemic foundation. These generalizations rely on established causal theories: laws of nature or, as less demanding alternatives, law-bound explanations of causes and effects.[137] They categorize and explain the phenomena to which they refer through those theories' lenses. Unlike statistical generalizations, they never say that "things just happen." Causal generalizations explain *why* things happen as they do. This pivotal feature allows people to ascertain the applicability of those generalizations to their individual circumstances.[138]

The causative system of probability rejects each and every mathematical rule of probabilistic calculus. This system does not recognize unevidenced probabilities. Consequently, it has no room for the complementation principle.[139] Causative probability is a function of evidential support. The presence and extent of this support are strictly empirical matters. When this support is present, the underlying scenario becomes probable. The scenario's probability is a function of the support's size and scope. Because the reasoner never has full information, the size and scope may be incomplete. Their incompleteness, however, does not increase the probability of the opposite scenario. To be probable, this scenario needs to have evidential support of its own. If it does not have any evidential support, its causative probability willbe zero, which simply means absence of information upon which a person canbase her decision. By the same token, if a scenario's confirmatory evidence is scant, its probability should be assessed as low even when the probability of the opposite scenario is not high either. The conflicting sets of information need not add up to 100% (as they do under the complementation principle).

The multiplication principle for conjunctions (the product rule) also becomes inapplicable.[140] The rationale for that rule is obvious: a compound two-event gamble is riskier than a gamble on one of the two events. When a person tosses a fair coin once, his probability of getting heads equals 0.5. When he tosses it twice, his probability of getting two heads in a row goes

---

[134] *See, e.g.*, Laurence H. Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1330 n.2 (1971) (attesting that "*all* factual evidence is ultimately 'statistical,' . . . in the epistemological sense that no conclusion can ever be drawn from empirical data without some step of inductive inference").

[135] *See supra* note 62 and sources cited therein.

[136] *See* Thomson, *supra* note 28, at 127–33.

[137] *See supra* note 62 and sources cited therein.

[138] *See* STEIN, *supra* note 69, at 91–106.

[139] *See* COHEN, *supra* note 5, 157–59.

[140] *See id.*; *see also* COHEN, *supra* note 23, at 198 ("A proposition's inductive support on given evidence has nothing to do with mathematical probability.").

down to 0.25. The causative system of probability, by contrast, is not a system of gambling. This system's sole criterion for probability is evidential support. Under this system, the evidential support for scenario *A* does not shrink when the reasoner considers the occurrence of that scenario in combination with scenario *B*. The evidential support for scenario *B* will not fade away either. Instead, the weaker of the two supports will determine the probability of the scenarios' conjunctive occurrence. The epistemic strength of the inferential chain of *A* and *B* will thus be determined by its weakest link.[141]

For identical reasons, the causative system of probability renders the disjunction rule irrelevant as well. This rule holds that the mathematical probability of several alternate scenarios equals the weighted sum of those scenarios' individual probabilities—a mirror image of the multiplication principle. Under the mathematical system, when a person participates in a series of gambles, her probability of succeeding in *one* of those gambles increases with the number of gambles. The causative system of probability employs an altogether different logic: the logic of evidential support. Under this system, the fact that a person needs to be correct only once does not improve the evidential support of her alternate causative hypotheses. The person's task is to make the best factual determination under incomplete information, not to maximize her expected payoff from a series of gambles. Hence, the probability that one of the person's hypotheses is correct equals the highest probability that attaches to one of those hypotheses individually.

B. *Evidential Variety as a Basis for Inference*

The logical composition of the two systems of probability—mathematical, on the one hand, and causative, on the other—reveals the systems' relative strengths and weaknesses. The mathematical system is most suitable for decisions that implicate averages. Gambling is a paradigmatic example of those decisions. At the same time, this system employs relatively lax standards for identifying causes and effects. Moreover, it weakens the reasoner's epistemic grasp of her individual case by requiring her to abstract away from the case's specifics. This requirement is imposed by the system's epistemically unfounded rules that make individual cases look similar to each other despite the uniqueness of each case. On the positive side, however, the mathematical system allows a person to conceptualize her probabilistic assessments in the parsimonious and standardized language of numbers. This conceptual framework enables people to form and communicate their assessments of probabilities with great precision.

The causative system of probability is not suitable for gambling. It associates probability with the scope, or variety, of the evidence that confirms the underlying individual occurrence. The causative system also employs

[141] *See* COHEN, *supra* note 5, at 160–61 (providing a technical demonstration of how the lowest level of causative support determines the probability of the underlying hypothesis).

rigid standards for establishing causation. Correspondingly, it disavows instantial multiplicity as a basis for inferences and bans all other factual assumptions that do not have epistemic credentials. These features improve people's epistemic grasps of their individual cases. The causative system has a shortcoming: its unstructured and "noisy" taxonomy. This system instructs people to conceptualize their probability assessments in the ordinary day-to-day language. This conceptual apparatus is notoriously imprecise. The causative system therefore has developed no uniform metric for gradation of probabilities.[142]

On balance, the causative system outperforms mathematical probability in every area of factfinding for which it was designed. This system enables people to perform an epistemically superior causation analysis in both scientific and daily affairs. Application of the causative system also improves people's ability to predict and reconstruct specific events. The mathematical system, in contrast, is a great tool for understanding averages and distributions of multiple events. However, when it comes to an assessment of an individual event, the precision of its estimates of probability becomes illusory. The causative system consequently becomes decisively superior.

In the area of science, the most famous example of this superiority is Karl von Frisch's research of bees' behavior.[143] Von Frisch's research had established,*inter alia*, that bees discriminate between colors, shapes, odors, and tastes.[144] To prove that bees differentiate between colors, von Frisch attracted them to a transparent source of food: a piece of blue cardboard associable with sugar-water.[145] To eliminate the possibility that bees use a different clue, but are still color-blind, von Frisch attempted to attract them, simultaneously, to differently colored food containers. The bees still preferred the blue card over all others. To eliminate the possibility that bees recognize the blue card by its smell, von Frisch covered it with a plate of glass. To eliminate the possibility that bees recognize the blue card's location, von Frisch rearranged the cards in many different ways. In each of these experiments, the bees returned to the blue card. Finally, to eliminate the possibility that blue happens to be the only color that bees recognize, von Frisch experimented with all other colors to find out that colors that bees discriminate between include blue, blue-green, ultraviolet, and yellow. This research is an impeccable application of Bacon's scientific elimination

---

[142] For an attempt at developing formal language for causative probabilities, see COHEN, *supra* note 23, at 199–244.

[143] *See* KARL VON FRISCH, BEES: THEIR VISION, CHEMICAL SENSES AND LANGUAGE 4–34 (rev. ed. 1971). For a superb philosophical explanation of this research and its epistemological implications, see COHEN, *supra* note 23, at 129–35.

[144] *See* VON FRISCH, *supra* note 143, at 4–67.

[145] *Id.* at 5–10.

method and a perfect example of how the evidential-variety criterion works.[146]

People, of course, do not carry out such systematic experiments in ascertaining the facts that they need to know about in their daily affairs. There is no good reason for them to do so. All they need to do is to acquire as much information as they reasonably can and then follow the logic of causative probability. PartIV.B below demonstrates that ordinary reasoners do exactly this: the causative system of probability aligns with common sense (indeed, common sense is the ultimate source of that probability). Before discussing this descriptive point, however, I need to complete the normative analysis of the issue by revisiting my examples of the mathematical system's failures.

The radiologist's case[147] is the first example to which I return. There, the radiologist's point-by-point examination of Peter's scan results generated evidential variety: a broad base of individuated causal indicators that eliminate the possibility of cancer in Peter's brain. On the other side of the scales, Peter finds random errors interchangeably committed by the radiologist and the MRI machine. The statistical rate of false negatives that Peter might worry about is 0.19.

Under the causative system, however, Peter should not worry about this rate because there are no causal facts that attach the 0.19 figure to him individually. On the other hand, each and every parameter of the radiologist's diagnosis of no-cancer is causally related to the individual condition of Peter's brain. Peter therefore should accept the radiologist's diagnosis as most probably correct. He should ignore the 0.19 figure completely because this "probability" has no causal credentials. This figure is causally irrelevant to Peter's case. What Peter should not write off completely is the possibility of error that attaches to what the radiologist said and wrote about his brain. But this possibility only means that the radiologist accidentally errs. Her diagnoses are not error-proof, but the causative probability of what she told Peter about his brain remains overwhelmingly high.

Things would have been different if the radiologist's rate of erroneous diagnoses were in high numbers. Peter would then have to check the effectiveness of the radiologist's methodology and her ability to work with that methodology (more realistically, he would have to obtain a second opinion). These factors would have had crucial significance for Peter's decision as to whether to undergo the brain surgery. Note, however, that the statistical figure prompting Peter's inquiry would still be causally insignificant.

The same analysis applies to my tax-audit example[148] and the fugitive case.[149] Application of the evidential-variety standard to the tax-audit ex-

146 *Id.* 4–67; *see also* COHEN, *supra* note 23, at 130–31.

147 *See supra* text accompanying notes 24–27.

148 *See supra* text accompanying notes 43–51.

149 *See* United States v. Elliott, 467 F.3d 688 (7th Cir. 2006).

ample leads to a straightforward conclusion: the firm's audit is a highly unlikely scenario. The firm's documents eliminate every possible suspicion of fraud, and there is no reason to doubt the credibility of those documents. The 10% audit rate that attaches to firms with high reported income is a causally irrelevant statistic. The reasoner will do well to ignore that statistic.

The fugitive case exemplifies a full epistemic separation between the event-specific causative information and general statistics. The two bodies of information are incommensurable. They give rise to mutually incompatible inferences that cannot be combined into a coherent whole.

Specifically, the causative information encompasses the effective steps that the defendant, Mr. Elliott, had taken to escape from the law. Those steps included borrowing another person's identity with that person's consent and full cooperation, cutting off ties to nearly everything he had in his previous life, and moving into inconspicuous retirement in Arizona. After fifteen years of fugitive life, something went wrong, and the FBI arrested the defendant. Ex ante, however, nothing went wrong for this defendant. After his settling in Arizona, causal indicators associated with his prospect of being apprehended by the FBI virtually did not exist. Causal indicators that were present at that time gave Mr. Elliott every reason to believe that he would never see prison, and those indicators stayed with him for more than a decade. Under those circumstances, no general rate of fugitive apprehension could make Mr. Elliott's prospect of remaining free less than practically certain. The court therefore ought to have increased his prison sentence more steeply than it did.

The upshot of this discussion for incentives theory is straightforward. People often need to evaluate the consequences of their individual actions under conditions of uncertainty. When a person wants to maximize the accuracy of those evaluations, she should use the causative probability system. Using the mathematical system will compromise the evaluations' accuracy. Causative probability therefore should replace mathematical probability as a main normative benchmark for formulating legal incentives.

IV. Policy Implications

Integration of causative probability into individuals' decisions has profound implications for legal policy. In what follows, I identify several of those by revisiting the core policy recommendations of two highly influential schools of thought: mainstream economic analysis of law and behavioral economics. Each of those schools bases its recommendations on the axiomatized view of probability. As I explained in the Introduction, this view recognizes only one system of probability: the mathematical system.

Mainstream economic analysis of law uses mathematical probability to determine two key elements of its policy recommendations. The first element is the probability of the enforcement of a legal rule that determines the magnitude of expected, as opposed to actual, penalties and rewards.[150] The second element is the probability that attaches to good and bad consequences that people's actions bring about.[151] This probability determines the magnitude of expected, as opposed to actual, harms and benefits brought about by those actions. Both elements play a crucial role in mainstream economic formulations of individuals' incentives.[152] Those formulations assume that a rational person relies upon mathematical probability in choosing between courses of action that affect her and her society's well-being. Those formulations also assume that a lawmaker can align private incentives with society's benefit by engineering expected penalties and rewards for individuals. Mainstream economists have developed two specific recommendations for that engineering. First, the lawmaker can bring the mathematical probability of enforcement up (or down) by ordering law enforcers to step up (or reduce)[153] their enforcement efforts. Second, the lawmaker can leave the law enforcement's probability as shefinds it and introduce an appropriate upward (or downward) adjustment in the magnitude of penalties and rewards.[154]

My preceding discussion has demonstrated that these assumptions are invalid. Mathematical probability affects a rational person's incentives only in a very limited set of circumstances. A rational person will rely on that probability only when she has no causal information pertaining to her situation and, consequently, has no choice but to gamble. Randomized law enforcement is a good example of this type of uninformed situation. For example, when a tax agency audits one firm out of ten and selects that firm at random, an individual firm can rationally assume that its probability of being audited equals 0.1. But law enforcement, as we all know it, is predominantly a causative, rather than randomized, phenomenon. When agencies enforce the law in a particular way, they normally have reasons for doing so. Those reasons, unlike lotteries, are not determined by mathematical averages. Instead, they are determined by specific instances of conduct to which lawenforcers react. By the same token, a person cannot rationally rely upon mathematical probability in assessing the harms and the benefits that an action she is planning to take might produce. Causal processes generating those harms and benefits are event-specific. When they repeat themselves in a particular way, a person might try to find a causal explanation for the repetition. However, she should not care about the repetitions'

---

150 *See* SHAVELL, *supra* note 10, at 177–78.

151 *See id.*; *see also id*. at 1 (attesting that descriptive economic analysis of individual behavior focuses upon rational actors who "maximize their expected utility").

152 *See generally* A. Mitchell Polinsky & Steven Shavell, *The Economic Theory of Public Enforcement of Law*, 38 J. ECON. LITERATURE 45 (2000) (discussing the economic backdrop of law enforcement).

153 *See generally* Richard A. Bierschbach & Alex Stein, *Overenforcement*, 93 GEO. L.J. 1743 (2005) (identifying instances of unavoidable overenforcement of the law and showing how it can be counteracted by procedural rules that make liability less likely).

154 *See* Becker, *supra* note 4.

statistical rate unless it comes close to 100%. High statistical correlations indicate the possible presence of causal connections, and the person should try to identify those connections. All other correlations are causatively meaningless, and a rational person should ignore them completely, unless—once again—she has no other choice but to gamble.

Causative probability, therefore, should take over most parts of the rational-choice domain. It ought to replace the mathematical system as a basis for analyzing and formulating individuals' incentives for action. Causative probability should also function as a primary tool for estimating the values of uncertain harms and benefits that individuals' activities produce. In section A below, I specify these policy recommendations and explain how to operationalize them.

Behavioral economists use mathematical probability as a benchmark for rationality.[155] Based on this axiom, they conducted a series of experimental studies showing that ordinary people systematically misjudge probabilities.[156] According to those studies, people's probabilistic errors fall into well-defined decisional patterns (identified as "heuristics")[157] that violate the basic rules of mathematical probability. Behavioral economists claim that a person who commits those errors makes bad decisions.[158] Those decisions are detrimental to the person's own well-being and to the well-being of other people.[159] This diagnosis of bounded rationality calls for the introduction of state-sponsored paternalistic measures, ranging from "soft" to "hard."[160] Behavioral economists argue that the state should regulate people's choices whenever those choices depend upon probability.[161]

But what if the probability that ordinary people use is causative rather than mathematical? Behavioral economists uniformly ignore this possibili-

[155] *See infra* text accompanying notes 183–184.

[156] *See infra* text accompanying notes 183–214.

[157] *See generally* Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, *supra* note 13, at 3, 3–18 (describing three heuristics used to gauge probability and make predictions).

[158] Daniel Kahneman et al., *Preface* to JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, *supra* note 13, at xi, xi–xii ("With the introduction of Bayesian ideas into psychological research . . . psychologists were offered for the first time a fully articulated model of optimal performance under uncertainty, with which human judgments could be compared. The matching of human judgments to normative models . . . led to concerns with the biases to which inductive inferences are prone and the methods that could be used to correct them.").

[159] *See, e.g.*, Sunstein, *supra* note 16, at 74 (observing that "people seem to treat situations as 'safe' or 'unsafe,' without seeing that the real question is the likelihood of harm" and providing examples); Eric Johnson et al., *Framing, Probability Distortions, and Insurance Decisions*, 7 J. RISK & UNCERTAINTY 35, 39–40 (1993) (reporting experiments in which wrong probabilistic decisions led people to choose inferior hospitalization policies).

[160] *See* sources cited *supra* note 18.

[161] *See* sources cited *supra* note 18.

ty.[162] This ignorance makes their theories incomplete and possibly flawed as well.

As an initial observation, the reader needs to notice that the commonsense reasoning that people use in their daily affairs more or less completely aligns with the causative system of probability.[163] Behavioral economists who have found that this reasoning systematically fails to meet the standards of the mathematical system should therefore have moved their experiments to the side of causative probability. This move has never occurred for a simple reason: the axiomatized view, to which behavioral economists subscribe, does not recognize that there is such a thing as causative probability.[164] This exclusion has a far-reaching consequence: the branding of people who base their decisions upon causative probabilities as irrational (or as boundedly rational). As I will show below, this branding is unjustified.

Failure to recognize causative probability and investigate its uses does not merely make behavioral theories incomplete. Another consequence of this failure is the behavioral economists' inability to see the presence of causative probability in their own experiments. This omission undermines the experiments' validity. In section B below, I demonstrate that it foils some of the foundational experiments that define the field of behavioral economics.

A. Causative Probability and the Economic Analysis of Law

As I already explained, mainstream economic theory adopts mathematical probability in all of its models and policy recommendations.[165] Examples of this unqualified adoption are abundant. The most recent ones can be found in the academic discussions of tax evasion[166] and corporate fraud.[167]

---

[162] For an illuminating integration of causative and statistical modes of reasoning, see Tevye R. Krynski & Joshua B. Tenenbaum, *The Role of Causality in Judgment Under Uncertainty*, 136 J. EXPERIMENTAL PSYCHOL. 430 (2007) (demonstrating experimentally that people make generally correct statistical decisions when statistics they are asked to consider represent clear causal structures).

[163] Adjudicative fact-finding is a striking example of this phenomenon. *See* Ronald J. Allen, *The Nature of Juridical Proof*, 13 CARDOZO L. REV. 373 (1991) (attesting that adjudicative fact-finding is organized into narratives in which one event brings about another event); Michael S. Pardo & Ronald J. Allen, *Juridical Proof and the Best Explanation*, 27 LAW & PHIL. 223 (2008) (explaining the adjudicative fact-finding process as a quest for the best explanation of the relevant causes and effects). *Cf.* STEIN, *supra* note 69, at 48, 204–07 (arguing that a legal system's choice between causative and mathematical modes of probabilistic reasoning depends on its prior determination of how to allocate risk of error).

[164] For a partial recognition of causative probability by behavioral theorists, see Krynski & Tenenbaum, *supra* note 162.

[165] *See, e.g.*, POSNER, *supra* note 3, at 11 (defining expected gains and losses by reference to mathematical probability); SHAVELL, *supra* note 10, at 4 ("[E]conomic analysis emphasizes the use of stylized models and of statistical, empirical tests of theory . . . .").

[166] *See, e.g.*, Lawsky, *supra* note 43, at 1041–57 (recommending a shift to the subjectivist version of mathematical probability and the corresponding determination of expected penalties for tax evasion by taxpayers' degrees of belief); Alex Raskolnikov, *Crime and Punishment in Taxation: Deceit, Deter-*

These discussions assume without argument that mathematical probability is the right tool—indeed, the only tool—for estimating individuals' expected gains, harms, penalties, and rewards under conditions of uncertainty. Based on this assumption, the discussions formulate their different proposals for law enforcement. Alas, in matters of law enforcement and in all other causative affairs, mathematical probability misses the target. This probability's core criterion—instantial multiplicity—moves reasoners away from the individual causes and effects that should be the basis of their decisions. Instead of focusing upon those causes and effects, reasoners are told to base their decisions on mathematical averages, either observed[168] or intuited.[169] This averaging weakens the reasoners' epistemic grasps of their individual situations, relative to what they could achieve under the causative system of probability.

The proposed switch from the mathematical system to causative probability has a good real-world illustration: the Supreme Court of Texas decision in *Sun Exploration & Production Co. v. Jackson*.[170] This decision was about ranch owners who gave an oil, gas, and mineral lease to an oil exploration and production company. The lease covered 10,000 acres of the owners' land and incorporated the company's implied covenant to reasonably develop and explore the land.[171] The dispute between the owners and the company concerned the owners' allegation that the company breached this covenant by failing to search for new oil and gas on their land. Based on this allegation, the owners attempted to cancel the lease. They substantiated this allegation by referring to the statistical chances of finding new oil and gas on their land and a high expected value deriving therefrom.[172] The owners argued that this expected value exceeded the company's exploration costs and, as such, activated the company's "exploration and development" duty.

The company disagreed with this statistical understanding of the "exploration and development" covenant. This covenant, it argued, could only be activated by a "known and producing formation"[173]—a concrete causal

---

rence, and the Self-Adjusting Penalty, 106 COLUM. L. REV. 569, 576–77 (2006) [hereinafter Raskolnikov, *Crime and Punishment*] (using the frequentist version of mathematical probability to determine expected penalties for tax evasion); Alex Raskolnikov, *Revealing Choices: Using Taxpayer Choice to Target Tax Enforcement*, 109 COLUM. L. REV. 689, 715–17 (2009) (same).

[167] *See, e.g.*, Miriam H. Baer, *Linkage and the Deterrence of Corporate Fraud*, 94 VA. L. REV. 1295, 1335–38 (2008) (using mathematical probability to determine expected benefits from corporate fraud and to mete out optimal penalties).

[168] *See* Raskolnikov, *Crime and Punishment*, *supra* note 166, at 576–77.

[169] *See* Lawsky, *supra* note 43, at 1041–44.

[170] No. C-6000, 1988 WL 220582 (Tex. July 13, 1988). After re-argument, the court replaced this decision with a different opinion in *Sun Exploration and Production Co. v. Jackson*, 783 S.W.2d 202 (Tex. 1989). Consistent with my present discussion, this opinion affirmed the previous finding that there was no breach of the covenant to reasonably develop an oil, gas, and mineral lease. *Id*. at 205.

[171] *Sun Exploration*, 1988 WL 220582, at *6.

[172] *Id.* at *10 (noting the owner's reliance on the "expected value" test).

[173] *Id.* at *7.

indicator of the presence of oil or gas deposits on the owners' land. This argument rejected the owners' claim that a naked statistical expectation can also activate the covenant. The company thus argued that its duty to develop and explore the owners' land could only be based upon causative probability of finding oil or gas.

The Supreme Court of Texas agreed with the company:

> Notwithstanding the evidence of a positive expected value on the prospects, a 6 1/2 percent, 8 percent, or even 25 percent chance of discovering hydrocarbons from on any given well does not provide a reasonable expectation of profit such that a court should force a lessee to drill or lose the lease. It is but mere speculation which operators and lessees occasionally assume in hopes of great profit. But, it is not sufficient proof for a court to force an unwilling operator to drill.[174]

This approach should apply across the board. To operationalize it, economic analysis must develop an appropriate substitute for its statistical expected-value methodology. To this end, it must devise a viable method of valuating individuals' prospects on the basis of causative probability. This method is available. The value of a person's welfare-increasing prospect equals the full value of that prospect minus the amount that the person would pay to make the welfare-increase certain. By the same token, the negative value of a person's welfare-decreasing prospect equals the amount that she would pay for the prospect's elimination. The amount that the prospect's holder would pay for the elimination of the undesirable uncertainty thus determines the uncertainty-discount attaching to the prospect.

This discount will vary from one case to another. The discount's size will depend on two factors. One of those factors is the scope of the evidence that confirms the person's favorable and unfavorable scenarios. Another factor is the person's disposition toward risk and uncertainty. This valuation system is similar to the conventional expected-value methodology in every respect except one. This system will use case-specific evidence of causation instead of statistics. Under this system, valuation of a person's prospect will rely exclusively upon causes and effects that are both individuated and empirically confirmed. The person will exercise her judgment to determine the dependability of the relevant causal information and the risk of error she is willing to tolerate. In making that determination, the person will ignore all information that has no causative impact on her individual case. As a general rule, she will disregard all unevidenced scenarios, including those that are statistically possible. The person should ignore the statistical figures indicating the general likelihoods of those scenarios. She may need those figures only for appraising the completeness and the consequent dependability of her causal information.

Under this approach, a person's appraisal of the uncertainty-discount will often be rough and intuitive—yet this is not a cause for concern. For

[174] *Id.* at *11.

reasons I already provided, causative probability allows a person to develop a better epistemic grasp of her individual situation than the one she would have under the mathematical system. A rational person should make full use of her reasoning tools to develop this better grasp. Those tools include the person's intuition and common sense. The inevitable imprecision of the person's intuitive evaluations does not make those evaluations inaccurate or unreliable. As far as the person's individual case is concerned, those evaluations are more accurate than statistical averages. The person, therefore, should rely on those evaluations.

The difference between the causative system of prospect-valuation and the conventional expected-value methodology can be substantial. The well-known problem of defensive medicine illustrates this difference. Medical malpractice law requires a doctor to inform her patient about every significant risk associated with the patient's condition and treatment.[175] The doctor also must identify and eliminate any such risk to the extent feasible.[176] The prevalent understanding of those rules associates the risk's significance with expected value: even when the risk of an adverse consequence to the patient is small, it willstill qualify as "substantial" if the consequence is death or serious injury.[177] This understanding of the law motivates doctors to shield themselves against malpractice suits by diagnosing patients for prospects of illness that are purely statistical.[178] Doctors also inform patients about those statistical possibilities and often recommend costly procedures and preventive measures to eliminate them.[179]

These departures from the causative determination of the patient's individual needs are epistemically unjustified. As far as health policy is concerned, they are also wasteful and morally deplorable. Under the causative system, in contrast, the value of an evidentially unconfirmed prospect of harm equals zero. This means that a doctor would not be required to eliminate such prospects by expending her valuable time and efforts. Nor would

---

175 See, for example, the oft-cited case *Canterbury v. Spence*, 464 F.2d 772 (D.C. Cir. 1972).

176 *See* DAN B. DOBBS, THE LAW OF TORTS § 242, at 633 (2000) (articulating doctors' duty to treat patients with customary reasonable care).

177 *See Canterbury*, 464 F.2d at 794 (holding that a 1% risk of paralysis falls within the spectrum of doctors' disclosure obligations).

178 *See, e.g.*, Sherman Elias et al., *Carrier Screening for Cystic Fibrosis: A Case Study in Setting Standards of Medical Practice, in* GENE MAPPING: USING LAW AND ETHICS AS GUIDES 186 (George J. Annas & Sherman Elias eds., 1992) (attesting that doctors urge prenatal patients to undergo comprehensive genetic tests to diagnose fetal abnormalities on the basis of statistics without considering individual medical needs).

179 *Id.* These procedures include unnecessary diagnoses, hospitalizations and referrals to specialists, needless gathering of laboratory information, and prescriptions for unneeded medications. *See* MASS. MED. SOC'Y, INVESTIGATION OF DEFENSIVE MEDICINE IN MASSACHUSETTS (2008), *available at* http://www.massmed.org/AM/Template.cfm?Section=Research_Reports_and_Studies2&TEMPLATE=/CM/ContentDisplay.cfm&CONTENTID=27797 (specifying those procedures, referencing empirical studies estimating nationwide annual cost of defensive medicine at between $100 billion and $124 billion, and calculating that Massachusetts alone spends about $1.4 billion on defensive medicine).

she have to discuss such prospects with her patients. The proposed reform consequently would reduce the volume of defensive medicine.

The conventional expected-value methodology, however, should not be abandoned completely. This methodology should apply in evaluating statistical risks of harm whenever those risks are actionable in torts.[180] Courts also should continue using this methodology in appraising the value of people's earning prospects when the dilution of those prospects constitutes compensable damage. Finally, the expected-value methodology is suitable for determining the scope of liability for recurrent torts,[181] and one can think of other examples as well.[182] This methodology, however, should only apply in well-defined areas of the law that call for statistical appraisals. Courts and lawmakers should not use it as a norm.

B. Causative Probability and Behavioral Economics

Introduction of causative probability into mainstream economic theories of the law brings about enrichment and refinement. Theories that account for causative probability expand their ability to explain social phenomena and make welfare-improving recommendations. In the domain of behavioral economics, the parallel consequence for many experimental models is unraveling.

Consider the field's flagship experiment, widely known as "Blue Cab."[183] The experimenters informed participants about a car accident that occurred in a city in which 85% of cabs were Green and the remaining 15% were Blue. The participants also heard a witness testify that the cab involved in the accident was Blue. The experimenters told the participants that this witness correctly identifies cabs' colors in 80 out of 100 cases. Based on this evidence, most participants decided that the probability of the victim's case against the Blue Cab Company equals 0.8.[184] This estimation aligned with the given credibility of the witness but not with the basic rules of mathematical probability. Under those rules, the prior odds attaching to the scenario in which the cab involved in the accident was Blue rather than Green—$P(B)/P(G)$—equaled 0.15/0.85. To calculate the posterior odds—$P(B|W)/P(G|W)$, with W denoting the witness's testimony—these prior odds had to be multiplied by the likelihood ratio.[185] This ratio had to be deter-

[180] *See, e.g.*, ARIEL PORAT & ALEX STEIN, TORT LIABILITY UNDER UNCERTAINTY 101–29 (2001) (identifying cases in which risks of harm should be actionable in torts); Ariel Porat & Alex Stein, *Liability for Future Harm*, *in* PERSPECTIVES ON CAUSATION (Richard S. Goldberg ed., forthcoming 2011), *available at* http://ssrn.com/abstract=1457362 (identifying conditions under which courts should impose tort liability for a wrongful creation of risk of future illness).

[181] *See* Saul Levmore, *Probabilistic Recoveries, Restitution, and Recurring Wrongs*, 19 J. LEGAL STUD. 691 (1990).

[182] *See* PORAT & STEIN, *supra* note 180, at 130–59 (using expected value as a basis for defining market-share liability and other forms of collective liability in torts).

[183] *See* Tversky & Kahneman, *supra* note 13, at 156–57.

[184] *Id.* at 157.

[185] For explanation of the "likelihood ratio" concept, see *supra* notes 40–42 and accompanying text.

mined by the odds attaching to the scenario in which the witness identified the cab's color correctly, rather than incorrectly: $P(W|B)/P(W|G)$. The posterior odds consequently equaled $(0.15 \times 0.8)/(0.85 \times 0.2)$, that is, 12/17. The probability of the victim's allegation against the Blue Cab Company thus equaled $12/(17 + 12)$, that is, 0.41. This probability falls short of the 0.5 threshold set by the "preponderance of the evidence standard" that applies in civil litigation. The 0.8 probability that most participants ascribed to the victim's case thus appears to be irrational.

But is it irrational? The participants were asked to combine together three items of information. Two of those items—the percentage of blue cabs in the city and the witness's rate of accuracy—were statistical. The third item—the witness's testimony "The accident involved a blue cab: I saw it"—was causative. This item of information was about the accident's effect on what the witness perceived, memorized, and reported. The individual cause-and-effect scenario captured by that item had nothing to do with the general statistics pertaining to cab colors and testimonial accuracy. If so, the participants' appraisal of the witness's credibility was not only rational from any plausible viewpoint but also more accurate than the experimenters' assessment.[186]

The participants in that experiment were asked to use the mathematical language. For that reason, most of them opined that the witness's credibility equaled 0.8 (or 80%).[187] Undoubtedly, those participants would also be happy to attest in words that the witness's testimony is most probably true. They evaluated the credibility of this testimony in terms of causative, as opposed to mathematical, probability.[188]

Note that the experimenters did not tell the participants that the distribution of blue and green cabs may have somehow affected the witness's capacity to tell blue from green. This causal connection would have been rather unusual, if not completely outlandish. The experimenters, therefore, should have informed the participants about that connection if they expected them to combine the three items into a single mathematical probability. The participants, for their part, rightly proceeded on the assumption that the distribution of blue and green cabs in the city was causally irrelevant to the witness's ability to identify colors correctly.[189] They also rightly discriminated between the causative and the statistical items of information. Those items are incommensurable: integrating them into a uniform statistical figure is an epistemological mistake. Most importantly, the participants correctly preferred the causative probability of the testimony's correctness over the mathematical one.

---

[186] *See* Alex Stein, *A Liberal Challenge to Behavioral Economics: The Case of Probability*, 2 N.Y.U. J.L. & LIBERTY 531, 536–37 (2007).

[187] *Id.*

[188] *Id.*

[189] *Id.*; *see also* COHEN, *supra* note 62, at 165–68 (explaining why it is rational for people to prefer causative probabilities, described as "counterfactualizable," to naked statistics).

The experimenters, however, did not recognize causative probability. Correspondingly, they did not distinguish between causative and statistical information from the beginning. Failure to discriminate between these two types of information is an epistemological error. As I already explained, statistical information is not extendible: it attests to the distribution of outcomes in a given sample of cases, but gives no reasons (other than instantial multiplicity) for predicting the occurrence of one outcome as opposed to another. The fact that 85 out of 100 cabs in the city are green has limited meaning: any witness will predominantly see green cabs on the streets of that city. This statistical fact, however, does not affect the operation of a witness's cognitive apparatus. When a witness sees a blue, a red, or another non-green car, she will normally recognize its color. This regularity is what the participants evidently proceeded upon in their decisions.

Most importantly, this regularity was law-bound rather than accidental: it referred to the witness's perception, memorization, and narration processes. This causative regularity consequently was extendible. Indeed, it was the only extendible piece of information that the experimenters asked the participants to consider. Under this set of facts, the participants' decision to evaluate the witness's testimony solely on the basis of this information was correct.

Behavioral economists do not seem to be aware of this insight. Based on the "Blue Cab" and similar experiments, they report a major finding: people systematically ignore prior probabilities.[190] Their next step is to find explanations for this systematic cognitive quirk. As it turns out, those explanations exist. According to behavioral economists, people use different "heuristics" instead of applying the Bayes Theorem.[191] The most prevalent of those heuristics are "representativeness"—a decisional shortcut that substitutes mathematical probability with the degree of personal familiarity or resemblance[192]—and "availability"—assessment of probability by "the ease with which instances or occurrences can be brought to mind."[193]

Begin with representativeness. Behavioral economists have developed a prototype experiment by which to identify this phenomenon. In that experiment, the experimenters ask participants to consider a person described by his neighbor in following words: "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail."[194] The experimenters then ask the participants to estimate the probability that Steve is engaged in a particular occupation that appears on their

---

[190] *See* Tversky & Kahneman, *supra* note 13, at 4–7, 156–58.

[191] *See generally* Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCIENCE 1124 (1974) (describing three different heuristics which are used to assess probabilities).

[192] *Id.* at 1124–27.

[193] *Id.* at 1127.

[194] *Id.* at 1124 (internal quotation marks omitted).

list. This list includes "farmer, salesman, airline pilot, librarian, or physician."[195] Most participants in this category of experiments ascribe high probability to Steve being a librarian.[196] They rely on familiar stereotypes but ignore the fact that farmers vastly outnumber librarians in the general population.[197] The participants, in other words, fail to account for Steve's low prior probability of being a librarian as opposed to farmer.[198] Arguably, this failure makes their decisions irrational, at least according to behavioral psychologists.

Under the causative system, however, it would be perfectly rational for a person to estimate that Steve is most likely to be a librarian. The general distribution of professions across population is causatively irrelevant to Steve's individual choice of occupation. His personality traits, in contrast, are causatively relevant to that choice (in a crucial way). Those factors consequently override the statistical numbers that attach to different occupations.

The same holds true for the derivative experiment featuring a thirty-year-old man, Dick, described to participants as married, but childless, and as a person with high ability and motivation who is liked by his colleagues and "promises to be quite successful in his field."[199] This description conveys no information whatsoever as to whether Dick is an engineer or an attorney. The group of professionals from which the experimenters drew this description included seventyengineers and thirtyattorneys,[200] and the participants knew it.[201] Under the mathematical system, therefore, the participants were supposed to report back that Dick has a 70% chance of being an engineer and a 30% chance of being an attorney. The participants, however, almost uniformly failed to give the experimenters this right statistical response. Instead, they assessed Dick's probability of being an engineer at 0.5.[202]

The experimenters, once again, have combined statistical information (the distribution of attorneys and engineers in the relevant sample) with causative factors (Dick's personality and choice of occupation). The inclusion of causative factors makes it completely plausible that the participants perceived their task as evaluation of case-specific evidence pertaining to Dick's choice of occupation. They may have evaluated Dick's *causative* probability of being an engineer rather than an attorney, or vice versa. This probability could only be extracted from the information relevant to Dick's occupational preferences.

---

195 *Id.*
196 *Id.* at 1124–25.
197 *Id.* at 1124.
198 *Id.* at 1124–25.
199 *Id.* at 1125.
200 *Id.*
201 *Id.* at 1124.
202 *Id.* at 1125.

The causative probability of Dick being an engineer was completely unknown. The participants' task, nonetheless, was to assess this unknown probability in numerical terms. Consequently, the participants must have analogized this task to a toss of an unrigged coin. This analogy explains their numerical assessment of the probability at 0.5.

Experiments confirming the presence of the availability heuristic include a recent series of studies of how people perceive risks of floods.[203] Two of those studies have found a strong correlation between the duration and intensity of a person's exposure to information about those risks and her estimation of how probable floods are.[204] A third study has demonstrated that a person's own experience with floods affects her determination of floods' probability.[205] These studies support the availability thesis because the participants' assessments of probability were completely unrelated to the statistical risks of flood.

This support is questionable. The first two studies involved participants with an imbalanced exposure to the "affect" information[206] and no prior knowledge of the floods' real probability. This informational imbalance was a product of the experimenters' manipulation. This manipulation created a compelling evidential variety that confirmed the pervasiveness of floods. Under the causative system, it is perfectly rational for a person to increase an event's probability on the basis of such evidence. The participants' failure to notice the experimenters' evidential manipulation is an altogether separate issue. This failure may be indicative of some reasoning defect, but it does not establish the presence of the "availability bias."

The third study involved good causative evidence: the participants' personal experience with floods. Absent evidence to the contrary, those participants were epistemically entitled to assume that their experience is no different from that of an ordinary person. The participants' failure to look for the general flood statistics, about which they have been asked, is indicative of a reasoning defect. The participants substituted what was supposed to be their assessment of a general risk with a causative probability determination. They evidently misunderstood their task, but this error involved no availability bias either.

Behavioral economists have recently identified another cognitive phenomenon: people's systematic differentiation between two types of uncertainty. One of those types relates to the meaning of a legal rule and another to whether the rule's violation will be punished.[207] This differentiation is responsible for people's unequal treatment of two probabilities: the proba-

---

[203] *See* Carmen Keller, Michael Siegrist & Heinz Gutscher, *The Role of the Affect and Availability Heuristics in Risk Communication*, 26 RISK ANALYSIS 631 (2006).

[204] *Id*. at 633–36.

[205] *Id*. at 636–37.

[206] *Id.* at 633–36.

[207] *See* Yuval Feldman & Doron Teichman, *Are All Legal Probabilities Created Equal?*, 84 N.Y.U. L. REV. 980 (2009).

bility that a particular conduct is punishable as a matter of law and the probability that a person acting in an unquestionably unlawful way will be punished as a matter of fact.[208]

As a general matter, experiments have shown that people heed the second probability more than the first: with the two probabilities being equal, people's rate of compliance with the law is higher in uncertain-enforcement situations than in cases of legal ambiguity.[209] Under the mainstream economic theory, the two probabilities are completely fungible: their effect on a rational person's expected punishment should be exactly the same.[210] Behavioral experiments have thus refuted the economic fungibility theory.[211] People participating in those experiments tended not to take advantage of the enforcement's low probability by committing an unequivocal violation of the law.[212] Their decisions followed a more refined pattern. On the one hand,"uncertainty stemming from the content of the law" motivated the participants "to perceive their acts as legal and therefore worthy (or at least not blameworthy)."[213] On the other hand,"uncertainty stemming only from the likelihood of enforcement—in situations in which the illegality of an action [wa]s clear"—made them "view the behavior itself as wrong."[214]

Failure to account for causative probability makes this insightful behavioral analysis incomplete. The meanings of ambiguous legal rules are not determined at random. Rather, courts determine those meanings by applying precedent, custom, analogy, policy analysis, and formal rules of interpretation.[215] Courts' application of these reasoning methods is regular and uniform.[216] These features make legal interpretation a causal phenomenon. Correspondingly, the probability that a court will prefer one interpretation of a legal rule over another is causative, rather than statistical.

The experiment's participants were asked to analyze a case of a possibly unlawful pollution and received the following description of "legal uncertainty":

> [T]he questionable action (disposing of the chemical into the lake) may or may not be deemed illegal because the chemical is relatively new and its legal status has not yet been determined; if the action is illegal, however, enforcement

---

[208] *Id.* at 991–97.
[209] *Id.* at 1009–11.
[210] *Id.* at 986–91.
[211] *Id.* at 997–1009.
[212] *Id.*
[213] *Id.* at 985.
[214] *Id.*
[215] *See generally* RICHARD A. POSNER, HOW JUDGES THINK (2008) (discussing theories of judicial behavior).
[216] *Id.* at 230–48 (observing that judges' applications of the law are a mix of interpretive and predictable pragmatic considerations).

> is certain because the authorities will be able to identify the factory that poured the chemical into the lake.[217]

The experimenters informed the participants that, because of this legal indeterminacy,"the overall likelihood of punishment (the probability of illegality multiplied by the probability of successful prosecution) is ten percent."[218]

Unfortunately, neither "10%" nor any other statistical figure can capture the real—causative—probability of the event in which a court interprets the rule in question in a way that makes the pollution illegal. The court will not determine the rule's meaning by flipping a coin, by throwing a die or, more exotically, by inducing a monkey to choose between ten similar bananas, one of which carries the inscription "illegal." Rather, it will try to identify and evaluate the reasons that produce this interpretive result.

On the other hand, a rule's probability of being enforced can be both causative and statistical, depending on whether the enforcer randomizes its efforts. The experiment's participants seem to have received the rule's statistical probability of enforcement. The experimenters told them that "pouring the chemical into the lake is clearly illegal but that successful enforcement is unlikely as there is a low [10%] chance that the authorities would be able to detect the identity of the polluting factory."[219]

The participants' choices therefore implicated an intractable combination of causative and statistical probabilities. This mix makes it difficult to decipher the motive underlying the participants' inclination toward lawful behavior. Indeed, some of those participants may have been motivated by the desire to do the right thing. Others, however, may have emulated the self-interested choices of Holmes's "Bad Man" that rely upon causative probability.

CONCLUSION

Of the two determinants of economic value—utility and probability—the first occupies the forefront of law and economics scholarship, while the second stays in the background. Economically minded theorists of law continually scrutinize the concepts of utility and well-being, over which they disagree,[220] while assuming without discussion that there is only one ration-

---

[217] *See* Feldman & Teichman, *supra* note 207, at 998.

[218] *Id.*

[219] *Id.*

[220] These disagreements focus on the virtues and vices of the preference-based and objective conceptions of utility and well-being. *See, e.g.*, Matthew D. Adler & Eric A. Posner, *Rethinking Cost-Benefit Analysis*, 109 YALE L.J. 165, 197–204 (1999); Howard F. Chang, *A Liberal Theory of Social Welfare: Fairness, Utility, and the Pareto Principle*, 110 YALE L.J. 173 (2000); Louis Kaplow & Steven Shavell, *Fairness Versus Welfare*, 114 HARV. L. REV. 961, 1353–54 (2001); Martha C. Nussbaum, *Flawed Foundations: The Philosophical Critique of (a Particular Type of) Economics*, 64 U. CHI. L. REV. 1197 (1997); Scott Shapiro & Edward F. McClennan, *Law-and-Economics from a Philosophical Perspective*, *in* 2 THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 460 (Peter Newman

al system of probability. This system, so goes the assumption, is predicated on the mathematics of chance: a body of rules that derive factual data from instantial multiplicities.

This assumption is false. Both in daily affairs and in science, people make sustained efforts to distinguish causes and effects from coincidences and to identify causal laws upon which they can rely. Mathematical probability is fundamentally incompatible with this practice as well as with the fact that people generally perceive their physical and social environments as causal rather than stochastic. Lawmakers therefore should not be guided by the mathematical system in devising legal rules, nor should they set up rules interfering with individuals' decisions that fail to satisfy this system's demands. Both lawmakers and economically minded legal scholars should consider the introduction of the causative system of probability in place of its mathematical cousin. The causative system's criterion for assigning probability—evidential variety—clearly outperforms the mathematical rules that purport to create knowledge from ignorance and sacrifice empirical content for the sake of algebraic precision. Application of this criterion will substantially improve actors' abilities to analyze their individual prospects and risks and make better decisions concerning future outcomes. In many contexts, such as decisions about undertaking medical procedures, the improvement in the individual's decisionmaking process can save her life. In law, understanding that actors base their decisions and actions upon causative probability will lead to substantially improved policies and rules.

---

ed., 1998). For a nuanced and superbly balanced analysis of these disagreements, see Daphna Lewinsohn-Zamir, *The Objectivity of Well-Being and the Objectives of Property Law*, 78 N.Y.U. L. REV. 1669, 1675–1714 (2003). For philosophical analyses of utility and well-being, see JAMES GRIFFIN, WELL-BEING: ITS MEANING, MEASUREMENT AND MORAL IMPORTANCE (1986); SHELLY KAGAN, NORMATIVE ETHICS 29–41 (1998); MARTHA C. NUSSBAUM, WOMEN AND HUMAN DEVELOPMENT: THE CAPABILITIES APPROACH (2000); DEREK PARFIT, REASONS AND PERSONS 493–502 (1984); T. M. Scanlon, *Preference and Urgency*, 72 J. PHIL. 655 (1975); Amartya Sen, *Well-Being, Agency and Freedom: The Dewey Lectures 1984*, 82 J. PHIL. 169 (1985).