THE MEANING OF PROBABILITY JUDGMENTS: AN ESSAY ON THE USE AND MISUSE OF BEHAVIORAL ECONOMICS

*Charles Yablon**

In this essay, Professor Yablon challenges the assumption that behavioral heuristics—such as the availability heuristic—constitute "biases" that necessarily lead to errors in probability judgments. He notes that there are many different concepts of probability and, in many contexts, no agreed-upon method for determining the correctness of inconsistent probability judgments. Yet many legal academics and policymakers ignore these complex aspects of probability theory, assuming that statistical or frequentist probabilities are always to be preferred over subjective judgments of probability.

Professor Yablon argues that choosing between frequentist and subjective approaches to probability judgments in policymaking can only be done with great sensitivity to context and the quality of the information available. He provides examples of current policy debates where statistical probability is to be preferred, such as environmental protection, others where subjective probability judgments are preferable, such as settlement of litigation, and a third important category, including products liability, where neither approach is clearly superior.

In his *New York Times* column on June 25, 2001, noted political commentator William Safire offered an "early-morning line" on potential Democratic presidential nominees for 2004.[1] Al Gore led the pack, with odds of 2-1 on nabbing the nomination,[2] but many other candidates had a decent chance as well.[3] Safire gave Tom Daschle, John Kerry, and

* *Professor of Law, Benjamin N. Cardozo School of Law, Yeshiva University.*

Thanks to Alex Stein for enormously valuable comments and critiques of a prior draft of this piece. Thanks also to the Samuel and Ronnie Heyman Center on Corporate Governance at Cardozo for funding and support.

1. William Safire, *The Henny Poll*, N.Y. TIMES, June 25, 2001, at A17.
2. *Id.*
3. *Id.*

Chris Dodd each odds of 4-1.[4] Joe Biden and Joe Lieberman both had a 5-1 chance of the nomination.[5] Pat Leahy was at 6-1, Russell Feingold at 8-1, John Edwards at 9-1, and Richard Gephardt, the longest shot, was at 15-1.[6]

Even before the 2004 Democratic nomination had been decided, we could confidently state that Safire's judgments concerning the chances of these prospective nominees were erroneous. This is not because Safire's perceived frontrunner, Al Gore, later dropped out of the race. Indeed, Gore's subsequent withdrawal is in no way inconsistent with Safire's estimate that he had a thirty-three-percent probability of winning the nomination, and surely does not refute it. The problem was not with Safire's political savvy, but with his understanding of probabilities. Each of Safire's odds represents a probability statement, which, like all mathematical probability statements, can be expressed as a number between zero and one or a percentage between zero and one hundred.[7] Four to one odds means a 1/5 or twenty-percent chance of winning the nomination. Fifteen to one means 1/16 or 6.25%, and so on. Because there could be only one Democratic presidential nominee in 2004, the aggregate likelihood that one of these individuals would be the nominee could not exceed 100%. Safire's odds, when added together, exceed 168%.

Safire acknowledged his mistake in a second column,[8] and graciously cited the columnist who first pointed it out.[9] He then proceeded to restate the odds on Gore as 3-1.[10] Joe Lieberman was reduced to a 12-1 long shot and Joe Biden was "scratched."[11] No odds on the other potential candidates were given. Unlike Safire's first set of probability judgments, these numbers are unassailable, because they add up to less than 100%.[12] They are also completely unverifiable. Even our knowledge today of the results of the nomination neither proves nor disproves

4. *Id*.
5. *Id*.
6. *Id*.
7. This is sometimes referred to as "Pascalian" or mathematical probability. *See* L. Jonathan Cohen, *On the Psychology of Prediction*, 7 COGNITION 385, 385 (1979) [hereinafter Cohen, *Psychology of Prediction*].
8. William Safire, *Holmes's Horse's Dog*, N.Y. TIMES, Feb. 7, 2002, at A29.
9. *Id*. (citing Dan Seligman, *Why Journalists Can't Add*, FORBES, Jan. 21, 2002, at 66). As the title indicates, Seligman attributed Safire's mistake to a lack of mathematical skills, but I think that is a bit unfair. As this essay points out, it is the difficulty of the concept of probability itself, particularly when applied to individual events, that gives rise to many of these mistakes and "biases." *See, e.g.*, *infra* notes 14–18 and accompanying text.
10. Safire, *supra* note 8.
11. *Id*.
12. I am assuming here that Safire did not intend to leave his probability judgments with respect to the other seven potential candidates unchanged. If he did, then the total for those seven candidates, plus the new odds for Gore and Lieberman, would still exceed 100%. Rather, I interpret his revised prediction to be Gore: 25%, Lieberman: 7.69%, Biden: 0%, and all other potential candidates: 67.31%.

Safire's statement that Gore had a twenty-five-percent chance, while Lieberman's was only 7.69%.[13]

Safire's problematic prognostications illustrate two fundamental truths about probability judgments. First, that many people have a great deal of difficulty making consistent, coherent judgments using mathematical probabilities. Second, that probability judgments concerning unique future events can never be proved either true or false by subsequent events.[14] The frequent inability of people correctly to make and understand mathematical probability judgments has been widely recognized and has been the subject of much important psychological study in recent years.[15] It has also been a focus of the emerging discipline of behavioral economics, a field whose insights have been increasingly applied to law and policy studies.[16] Legal scholars now frequently discuss behavioral concepts like the "availability heuristic"—the tendency of most people to estimate the likelihood of the occurrence of a future event based on the ease with which they can recall similar past events.[17] These heuristics are often also described as "biases," which distort and interfere with people's ability to make "rational" judgments about risk and probability.[18]

What much of this work fails to appreciate, however, is the second lesson of the Safire story—that the accuracy of many probability judgments cannot easily be either proved or refuted. The meaning of probability judgments is controversial and complex.[19] Probability statements potentially invoke a number of different concepts of probability.[20] Those involving characteristics of groups, such as the percentage of students who will pass the bar exam, are generally amenable to objective empirical verification. Judgments on the probability of the occurrence of unique individual events like Safire's presidential picks, however, are, strictly speaking, not verifiable. No subsequent event can either establish

13. If we take Biden's "scratch" to mean he has no chance of the nomination whatsoever, then that is the only one of Safire's predictions that might have been disproved by subsequent events. Safire, however, could have readily fixed this problem by giving him a small, but nonzero, shot at the nomination, say .1%.

14. This assumes, of course, that the probability assigned is some fraction between one (certain to occur) and zero (certain not to occur).

15. *See* Russell B. Korobkin & Thomas S. Ulen, *Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics*, 88 CAL. L. REV. 1051, 1084–85 (2000).

16. *See id.* at 1053.

17. *See, e.g.*, Mark Seidenfeld, *Cognitive Loafing, Social Conformity, and Judicial Review of Agency Rulemaking*, 87 CORNELL L. REV. 486, 501 (2002).

18. Frances L. Edwards, *Worker Right-to-Know Laws: Ineffectiveness of Current Policy-Making and a Proposed Legislative Solution*, 15 B.C. ENVTL. AFF. L. REV. 1, 21 (1987) ("Biases are also found in the intuitive judgment of probability. These biases often interfere with rational decisionmaking."); Korobkin & Ulen, *supra* note 15, at 1085 ("Research in the behavioral sciences has demonstrated that individuals are systematically biased in their predictions of the probable results of various events."); Seidenfeld, *supra* note 17, at 501 ("The availability heuristic leads individuals to overestimate the probability of an event that comes easily to mind.").

19. *See infra* note 43 and accompanying text.

20. *See infra* note 43 and accompanying text.

or refute their accuracy. Yet this does not stop us from arguing about them and even citing evidence that we consider relevant to support our views.

Many difficult probability judgments relate to the likelihood of the occurrence of unique future events, like assessing the risk that a particular smoker will die from lung cancer. Although the event is a unique one and the accuracy of the probability therefore cannot be conclusively proved by subsequent events, we nonetheless believe that statistical data about cancer death rates among smokers is relevant to making an accurate assessment of that probability.[21] Yet it is also likely that certain smokers have available to them information about their own individual risks from smoking that could rationally support their belief that their own risk from smoking is significantly different from the statistical average. For example, they may have a particularly favorable family medical history, or they may already be seventy-five years old. Accordingly, if there is a disparity between the subjective probability judgment of an individual regarding a unique future event and the probability implied by the statistical data, we may, in certain contexts, view the statistically based probability judgment as the more accurate one. But we are not compelled, either by logic or empirical evidence, to such a conclusion.

The purpose of this essay is to demonstrate that when policymakers and law professors cite behavioral studies to "prove" that people systematically "overestimate" the probability of the occurrence of an uncertain event, or make "incorrect" assessments of some potential risk, their arguments are at best incomplete, and at worst incoherent.[22] It shows first that availability and other heuristics that people use to make probability judgments do not necessarily lead to "errors." Rather, the accuracy of such judgments must be evaluated on a case-by-case basis according to the type of probability involved and other relevant information available on the causal relationship that is being predicted. Second, it demonstrates that when the probability judgment involves prediction of unique future events, no empirical data can ever conclusively verify or refute the accuracy of such a judgment. Third, while statistical probability can sometimes be used to cast doubt on the accuracy of an individual's subjective probability judgments, this can never be based on broad generalizations from people's presumed "biases" or "heuristics." Rather, it requires a context-specific empirical inquiry and judgment with respect to the issues involved. Critical to this inquiry is whether the individual, in making a personal probability judgment, has available additional significant information that is relevant to the causal relationship being predicted and that is not reflected in the statistical data. In many cases, we may decide that such individual probability assessments, even if poten-

21. *See infra* text accompanying notes 256–67.
22. *See infra* Part II.

tially subject to heuristics such as availability, are more likely to be accurate than probabilities derived solely from statistical data.

The prime determinant of this decision will generally be our judgment of the relative amounts of data available from both statistical and individual sources that are relevant to the causal processes being predicted. Such judgments will generally reflect our scientific beliefs concerning the nature of the causal processes involved.[23] When we believe that the causal processes involved in the event we are predicting are uniform or stochastic and repetitive in nature, like those involving purely physical processes such as tire failures or roulette wheels, we are more likely to rely purely or primarily on statistical data. Conversely, when we believe the event being predicted will be the result of a relatively unique and unusual confluence of many nonrecurring factors or involving individual decision makers, such as the 2004 Democratic presidential nomination, we are more likely to put our faith in the subjective assessments of knowledgeable observers.

Furthermore, the quality of the relevant information must also be assessed. Even in areas where we assume the causal processes are relatively uniform and stochastic, such as disease etiology, we may prefer an expert medical opinion if the epidemiological data is poor or inconclusive. Conversely, we may doubt an individual's probability assessment of a unique future event if we believe the individual's judgment to be unreliable, or strong statistical evidence about similar events seems to make the prediction unlikely.[24]

Accordingly, in each case where behavioral theory is cited to impugn a subjective probability assessment, the specific probability statements involved must be carefully analyzed to determine the basis for the claim of "error." In a very few cases, the errors will be like that of Safire's first presidential odds, an internally inconsistent application of mathematical probabilities. More commonly, however, the claimed "errors" consist only of a potential disagreement between an individual's subjective probability assessment and probability assessments derived statistically from some previously analyzed historical data. It is a central

23. The relationship between questions of probability and of scientific induction has been recognized for some time. Robert Nozick describes the issue, and its historical source, as follows:

> The modern source of reliability views is Charles Pierce, who spoke of the validity of rules of inference ("leading principles") in terms of the percentage of times that, when their premises are true, their conclusions are true. Deductively valid rules score 100 percent, while valid inductive rules would show a very high percentage.

ROBERT NOZICK, THE NATURE OF RATIONALITY 66 (1993) (citing CHARLES S. PIERCE, *The Fixation of Belief, in* 5 COLLECTED PAPERS 223, 223–47 (1931–1958), *reprinted in* THE PHILOSOPHY OF PIERCE: SELECTED WRITINGS 5–22 (J. Buchler ed., 1950)); *see also* L. JONATHAN COHEN, AN INTRODUCTION TO THE PHILOSOPHY OF INDUCTION AND PROBABILITY §§ 2–4, at 4–39 (1989) [hereinafter COHEN, INDUCTION & PROBABILITY].

24. For example, we may doubt a teenager's statement "I'm going to be a rock star" even though, as a causal matter, individuals' statements of intent are usually very good predictors of their own future conduct. In this case, we view the statement as unreliable because we know the statistical odds against becoming a rock star are long.

point of this essay, however, that statistically derived frequentist probabilities are not necessarily more accurate than individual probability assessments. Indeed, in predicting the likelihood of unique future events, there are often good reasons to prefer subjective probability assessments to statistical frequentist data.

For example, if I refuse to use cell phones for fear of radiation damage, but fail to get the radon levels in my house checked, you might well call me irrational, in an empirical, although not a logical, sense.[25] Pretty good statistical evidence indicates that radon is a significantly greater risk for Americans generally, and because both risks are believed to be uniform and random throughout the population, I am unlikely to have good scientific grounds for believing that my individual risks are different.[26]

If, due to terrorism concerns, I choose to drive from New York to Washington rather than fly, you might disagree with the decision, but you would be unlikely to call it irrational because, particularly after September 11, one might rationally conclude that there is no good way to determine the true risk of flying. Moreover, even if statistics indicate that driving, is, on average, more dangerous, I might be a particularly safe and skillful driver.[27]

Finally, if I settled a medical malpractice lawsuit for forty cents on the dollar, without investigating the law or the facts, merely because I know that, statistically, only forty percent of all medical malpractice claims are successful,[28] you would call that decision crazy—and probably legal malpractice—despite, indeed, precisely *because* I adhered to strict frequentist concepts in a situation that appears to call for a more subjective and individualized probability assessment.

This essay will examine why each of these different approaches to probability assessment makes sense in the specific context in which it is made. It argues that similar context- and information-specific evaluations of frequentist and subjective approaches to probability assessment must be made whenever there is an effort to apply behavioral concepts, such as the availability heuristic, to policy issues. Availability and similar behavioral heuristics are not cognitive defects that necessarily lead people to inaccurate judgments. Rather, they are methods of dealing with

25. According to the website maintained by the Environmental Protection Agency, radon exposure is believed to be the second leading cause of lung cancer death in the United States, exceeded only by smoking. *See* ENVTL. PROT. AGENCY, RADIATION: RISKS AND REALITIES, *at* http://www.epa.gov/radiation/docs/risksandrealities/rrpage3.html (last visited June 7, 2004). The FDA's website on cell phones, in contrast, states: "The available scientific evidence does not show that any health problems are associated with using wireless phones. There is no proof, however, that wireless phones are absolutely safe." Food & Drug Admin., *Cell Phone Facts*, *at* http://www.fda.gov/cellphones/qa.html#22 (last visited Oct. 5, 2003).

26. *See supra* note 25.

27. However, my belief that I am a particularly safe and skillful driver may itself be false.

28. For a comparison of actual win rates in medical malpractice and other cases, see Kevin M. Clermont & Theodore Eisenberg, *Trial by Jury or Judge: Transcending Empiricism*, 77 CORNELL L. REV. 1124, 1136 (1990).

uncertainty, and may sometimes be the best methods available. Moreover, uncertainty frequently gives rise to a combination of responses, emotive as well as cognitive, such as optimism or fear. Viewing these as simply cognitive errors misses an important element of human response to risk and uncertainty.[29] Sometimes, optimism (or fear) may seem irrational (in a limited empirical way) in light of the best information we have about the problems facing us. At other times, however, it may be the best available response to uncertainty, and precisely the one we wish to encourage. The use of behavioral theory in law will be enriched if we recognize the critical role that context, theories of causation, and information quality play in making these determinations.

This essay explores the propriety of choosing frequentist or subjective probability assessments in many different contexts of interest to policymakers. Part I describes in some detail the concept of probability and the difficulty of applying probabilistic concepts to individual uncertain events.[30] It also differentiates and explains the concepts of frequentist[31] and subjective probability,[32] as well as other basic concepts and controversies in probability theory.[33] Part II focuses on Kahneman's and Tversky's classic studies in behavioral theory involving the availability heuristic.[34] It shows that Kahneman and Tversky, in their original article, recognized the limited applicability of those studies to predictions of individual future events.[35] They viewed their work as potentially enhancing an individual's ability to make subjective probability judgments, not as a demonstration that such judgments are inevitably flawed.[36] It then examines an important and relatively unknown challenge to those studies by the English philosopher L. Jonathan Cohen.[37]

The next three parts apply these theoretical considerations to invocations of the availability heuristic and probability judgments in recent legal scholarship.[38] These parts use the differing philosophical concepts of frequentist and subjective probability to analyze and distinguish policy debates involving probabilities where potential cognitive errors can be

29. Viewing the human response to uncertainty purely in terms of cognitive accuracy may not only be psychologically mistaken, but may also assume or ignore questions about the *utility* of predictions under uncertainty. As Robert Nozick points out, there may be times when the expected utility of believing something other than the "truth" may lead to greater happiness or utility than a cold-blooded probability assessment based on the best available evidence. NOZICK, *supra* note 23, at 69–70. For example, the widespread belief among people about to be married that their likelihood of divorce is well below fifty percent, *see infra* text accompanying notes 313–26, is cognitively inaccurate based on the best available objective data. Yet it is hard to see the advantages of holding the cognitively accurate belief that one's marriage has only a fifty-percent chance of success.

30. *See infra* Part I.

31. *See infra* Part I.A.

32. *See infra* Part I.B.

33. *See infra* Parts I.C–D.

34. *See infra* Parts II.A–B.

35. *See infra* Parts II.A–B.

36. *See infra* Parts II.A–B.

37. *See infra* Part II.C.

38. *See infra* Parts III–V.

fairly readily identified based on frequentist data from those where the claim that availability or other heuristics leads to cognitive errors in probability assessment is either controversial or simply unprovable.[39] Part III considers issues such as environmental policy, where aggregate risk is the prime consideration and frequentist approaches to probability assessment are generally appropriate.[40] Part IV considers issues of products liability and divorce policy, where questions of individual risk assessment are paramount, but where frequentist data about such risks may have a significant role to play.[41] Part V deals with issues relating to likelihood of contract performance and likelihood of success in litigation, where statistical data provides little or no basis for questioning the accuracy of subjective probability assessments.[42]

I. THE CONCEPTS OF PROBABILITY

Probability theorists recognize that probability is a concept that can be conceived of and understood in a variety of different ways.[43] While the mathematics of the Pascalian probability calculus can be clearly and consistently stated,[44] the meaning of these mathematical formulations, when applied to real world events, are subject to many different, almost contradictory, interpretations.[45] This lack of a single, consistent meaning for the term "probability" should not be all that troubling, or even surprising. Since Wittgenstein, we have become philosophically comfortable with the notion that words need not have a single, determinate meaning in all contexts.[46] Rather, like Wittgenstein's famous discussion of the word "game," the same word may denote quite a number of different things in different contexts.[47] Those various meanings may not be rigidly defined by a series of necessary and sufficient conditions. Rather, they may only be loosely linked by a "family resemblance," such that

39. *See infra* Parts III–V.
40. *See infra* Part III.
41. *See infra* Part IV.
42. *See infra* Part V.
43. While all agree that there are a number of different conceptions of probability, not all agree on what that number is. *See* JAMES BROOK, A LAWYER'S GUIDE TO PROBABILITY AND STATISTICS 33–40 (1990) (describing three theories of probability); COHEN, INDUCTION & PROBABILITY, *supra* note 23, at 43–80 (discussing six categories of probability theories); David H. Kaye, *Introduction: What Is Bayesianism?*, *in* PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE 1, 3–5 (Peter Tillers & Eric D. Green eds., 1988) (describing seven types of probability). An even bigger issue is whether Pascalian mathematical probability analysis (as well as Bayes's Theorem, which is an extension of it) provides the only normatively correct method of describing and analyzing probability judgments, or whether an alternative "Baconian" method of probability assessment is more appropriate in certain cases. *See infra* notes 86–101 and accompanying text.
44. *See, e.g.*, Cohen, *Psychology of Prediction, supra* note 7, at 388.
45. *See supra* note 43.
46. *See generally* L. WITTGENSTEIN, PHILOSOPHICAL INVESTIGATIONS (3d ed. 1958).
47. *Id.* ¶ 66.

baseball, roulette, and the FreeCell program on your computer are all properly described by competent speakers of English as "games."[48]

One purpose of this section is to demonstrate that "probability," like "game," does not name a single discrete concept, but rather a number of different concepts—notably frequentist and subjective concepts of probability, and possibly some others—which are also linked by a "family resemblance" to one another. A second purpose is to delineate these various concepts to arrive at an understanding of the way probability is used in recent work on behavioral theory and law. To do that, I want to consider two paradigmatic statements about probability that lawyers and policy analysts frequently make. One is the policymaker's statement that most people overestimate (or underestimate) the probability of some particular environmental risk.[49] The other is the lawyer's statement that there is a high or low probability of success with respect to a particular case.[50]

A. Frequentist or Statistical Probability

Consider first what is perhaps the most familiar concept of probability, the relative frequency or frequentist view.[51] Under this theory, a statement concerning the probability of a given event or outcome is a statement describing the relative frequency of that event or outcome within a given population.[52] For example, if there are fifty faculty members at my law school and ten of them smoke, the probability that any given member of the faculty is a smoker is 10/50 or twenty percent.

One important feature of this concept of probability is that it is empirically based and objectively verifiable. The statement that there is a twenty-percent probability that one of my colleagues is a smoker is true only if the ratio of smokers to total faculty is, as a matter of objective, empirical fact, twenty percent. But along with this empirical certainty comes a unique form of uncertainty. The probability only applies to the faculty as a group or to a randomly selected member of the faculty—someone about whom I have no specific information other than that they are a member of the faculty and that they have been "randomly selected." If I were told the faculty member's name, for example, then it is doubtful that I would still view the odds that he or she was a smoker as twenty percent. I might well know for a fact whether or not that person smoked. Thus, the frequentist concept of probability, when applied to individuals or individual events, generally involves both substantial information about the group to which the individual or individual event be-

48. *Id.* ¶ 67.

49. Ann Bostrom, *Risk Perceptions: "Experts" vs. "Lay People,"* 8 DUKE ENVTL. L. & POL'Y F. 101, 101 (1997).

50. Clermont & Eisenberg, *supra* note 28, at 1126–33.

51. *See, e.g.*, Kaye, *supra* note 43, at 3–4.

52. *See id.*

longs, as well as significant uncertainty or lack of information about facts relating specifically to that individual.

One important area where these features are frequently applicable is in the prediction of future events. We are unlikely to know many specifics about future events, but we can recognize them as belonging to a group about which we can develop significant statistical information. Statements concerning the risks or probabilities of various environmental or toxicological hazards almost invariably apply this frequentist concept of probability.

Consider, for example, Professor Sunstein's statement in the *Harvard Law Review*: "Do people know which risks lead to many deaths and which risks lead to few? They do not. In fact, they make huge blunders."[53] This statement only makes sense as an application of the frequentist concept of probability. Indeed, Sunstein's support for it is a study in which people were asked to compare two potentially fatal hazards and asked to estimate which was responsible for more deaths.[54] The question itself is frequentist, in that it asks for the relative frequency of various hazardous causes as a percentage of all fatalities. Such a question, at least with respect to past fatalities, has a single, potentially verifiable right answer. If people are asked about the likelihood that these hazards will cause fatalities generally, or to the average or random person, and their answers vary significantly from the objective probability, then they have indeed made a "blunder" and are laboring under a misapprehension of fact. This misapprehension about the cause of most fatalities may, in itself, have significant implications for policy analysis. It is not the same, however, as the claim that people are mistaken about their own individual risks of death from various hazards. That is a much more complex claim, which involves both the reasonableness of the assumption that the future will be like the past, and an analysis of the additional information which may be available to individuals when they evaluate their own personal risks.

For example, we can count, with a reasonable degree of certainty, the number of suicides and homicides that took place in recent years in the United States and say, with the same degree of certainty, that Americans were at greater risk of death from suicide than from homicide.[55] If we make the reasonable, but by no means necessary, assumption that suicide and homicide rates in the United States will be similar this year to

53. Cass R. Sunstein, *The Laws of Fear*, 115 HARV. L. REV. 1119, 1126 (2002) (reviewing PAUL SLOVIC, THE PERCEPTION OF RISK (2000)).

54. *Id.* at 1126–27.

55. According to data available on the website of the Center for Disease Control and Prevention, in 2000 (the most recent year for which data is available), there were 1.7 times more deaths from suicide in the United States than homicide. Ctr. for Disease Control & Prevention, *Suicide in the United States*, *at* http://www.cdc.gov/ncipc/factsheets/suifacts.htm (last visited June 7, 2004). Suicide was the eleventh leading cause of death for all Americans and the third leading cause of death for Americans aged fifteen to twenty-four. *Id.*

previous years' rates, we can also say, as an objective or empirical fact, that Americans as a group (or if you prefer, the "average American") have a greater probability of dying from suicide than from homicide.

If we then move, however, from the general (or average) to the specific, and ask if you, the person reading this essay, have a greater probability of dying this year from suicide or homicide, then it is unlikely your answer will be determined by statistical facts about homicide and suicide rates in the country at large. Rather, you will consider many personal facts about yourself, the neighborhood where you live, your history of mental health and perhaps that of your close relatives, and other past events in your life. You might well decide that your individual probability of suicide is less than the probability of homicide (although hopefully both are quite low). That would not be an objective description of any facts about the world, but could rather be viewed as a psychological description of your current beliefs, a subjective probability assessment that cannot be proved or disproved by any empirical evidence. Just because those beliefs are subjective, however, does not mean they are meaningless. They are subject to change through evidence and argument, and they have objective consequences in the actions you take, such as whether you go out at night in bad neighborhoods or take antidepressants. These issues will be discussed in more detail in part III of this essay.[56] For now, it is sufficient to note that the frequentist concept of probability underlies most discussion of risk and probability in such areas as environmental law and toxic torts, and provides an objective standard against which to demonstrate popular misconceptions.

B. Subjective or Personalist Probability

If we try to apply this same concept of probability to lawyers' predictions about the outcome of specific cases, however, we immediately run into problems. Frequentist probability may play some role in predicting legal outcomes, but it is surely not the predominant concept of probability being used. For example, when taking on a criminal appeal, a lawyer might caution the client that less than ten percent of such cases are reversed on appeal. We would not expect the lawyer to base her own evaluation of the likelihood of reversal solely on the relative frequency of such reversals among all criminal appeals. We would expect her to evaluate the likelihood of success based on all the individual factors relating to that specific case: the nature of the evidence presented at trial, the strength of the legal issues on appeal, the reputation of the trial judge, and the appellate court's prior rulings on similar issues. The frequentist view of probability gives us no understanding of how probability can be applied to the assessment of a single unique event. Unless an event can be categorized as one instance of a general set or class of

56. *See infra* Part III.

events, of which outcomes can be accurately measured and counted, the frequentist view of probability is inapplicable.[57] Yet individuals, and especially lawyers and judges, often make probability statements regarding the outcome of unique individual events, particularly lawsuits.[58] If such lawyers are not talking about relative frequencies, then what are they talking about?

Most probability theorists would say that such statements reflect a personalist or subjective concept of probability.[59] As the latter term implies, probability statements under this concept represent the degree of rational belief that the speaker holds about the likelihood of occurrence of the event in question.[60] Although such statements are not empirically verifiable in the way frequentist concepts are, they do have an objective correlative—the betting odds that the person holding the belief would be willing to give or accept as to the occurrence of the event. Accordingly, the difference between my saying that the Mets have a ten-percent chance to win the pennant and saying they have a fifty-percent chance is the degree of my subjective belief that the Mets will in fact win the pennant. If I will accept an even-money bet that the Mets will win, then I have a stronger belief than if I demand odds of 9-1.

The subjective concept of probability enables us to make sense of probability statements about individual, nonrepeat events, and provides a mechanism (i.e., betting behavior) for formulating and understanding such probability statements.[61] Utilizing the subjectivist concept of probability, we can also prescribe minimal standards of rationality with respect to beliefs about probability.[62]

57. A frequentist theory of probability is also unable to deal with infinite or unmeasurable sets. Because the computation of probability on a frequentist model requires determination of the numerical frequency of a given characteristic within a given reference class, the number of members of the reference class must be finite and measurable. The class of actually litigated cases (and its various subsets) would seem to be finite and measurable—and even if, as suggested above, lawyers do not refer to it often in making probability judgments, it is still a source of useful information about the probability of outcomes.

58. *See* Clermont & Eisenberg, *supra* note 28, at 1126–33.

59. See COHEN, INDUCTION & PROBABILITY, *supra* note 23, at 58–70.

60. *Id.* at 59.

61. The personalist concept of probability underlies much of modern decision theory. *See* DETLOF VON WINTERFELDT & WARD EDWARDS, DECISION ANALYSIS AND BEHAVIORAL RESEARCH 93 (1986). There have been some efforts to offer the tools of decision analysis to lawyers. *See, e.g.*, Kaye, *supra* note 43, at 10–13.

62. These rationality standards include avoiding beliefs that could result in a "Dutch book" being made against the holder of the beliefs. *See* Ian Hacking, *Slightly More Realistic Personal Probability*, *in* DECISION, PROBABILITY AND UTILITY 118, 131–32 (Peter Gärdenfors & Nils-Eric Sahlin eds., 1988). That is, the holder cannot have beliefs that would enable someone to bet against him for a sure net gain. For example, if I simultaneously believe there is a fifty-percent chance of the Mets, Cubs, and Pirates winning the pennant, and I took an even-money bet on each of the three, then I would (at best) lose two dollars for every one dollar won. This is the same error Safire made in his initial presidential picks. The "Dutch Book theorem" is extensively analyzed in the subjectivist literature. *Id.*

It is also argued that the use of Bayes's Theorem as a means of combining probability judgments will avoid the possibility of a Dutch Book, and therefore constitutes a rational method of reasoning with respect to probabilities. The extent to which Bayes's Theorem should be utilized by juries and

The subjectivist concept of probability, however, implies that an individual's subjective belief in an event's likelihood of occurrence is, in most instances, the only justification for any statement of probability. That is, if I assert that the Mets have a ninety-nine-percent likelihood of winning the pennant next year, then the statement is true insofar as I actually possess that degree of faith in the ultimate triumph of the Mets. The fact that everyone else thinks the statement highly implausible in no way refutes it, unless, of course, the knowledge that others disagree with me alters my subjective degree of belief. By the same token, the fact that two individuals disagree strongly about a probability judgment creates no inconsistency. Indeed, under a subjectivist view, it is hard to see what the disagreement is even about. If I assert that the Mets have a ninety-nine-percent chance of winning the pennant, and my colleague asserts it is a mere .5% chance (because the Cubs are the team of destiny in his view), then we are merely reporting on our different subjective states of belief, not disagreeing about any objective facts.

How well does the subjectivist concept of probability comport with the lawyer's activity of predicting future case outcomes? Quite well, at least in certain respects. It is certainly plausible to view many of the decisions lawyers make with respect to litigation and settlement as "bets," which reflect their degree of belief about the likelihood of the outcomes of particular, nonrepeat events. The subjectivist concept of probability captures that aspect of legal prediction and permits the application of a sophisticated mathematical analysis to the probabilities of single events.

What the subjectivist view does not provide is any justification, other than subjective belief and certain minimum rationality constraints, for the assignment of particular probabilities to particular events. If I believe that a case has a ten-percent likelihood of success, and you believe it has a seventy-percent chance, then the subjectivist concept gives us no way of resolving the dispute, or even of explaining why it is a dispute and not merely a disparity of beliefs. Yet, surely two lawyers who reached such disparate conclusions about the likelihood of success on a case would (1) view the simultaneous assertion by two trained lawyers that the same case had a ten-percent and a seventy-percent chance of success as somehow involving inconsistent judgments, and (2) seek to resolve the inconsistencies through discussion of objective features of the case. In short, they would act as if the probability were to be determined by an analysis of objective features of the case and that consensus as to a "correct" probability were obtainable through such analysis.

The subjectivist view of probability does not deny that people reach their individual probability assessments through an analysis of the objective facts surrounding the event.[63] Rather, it takes that as a brute fact of

other legal decision makers has been the subject of an interesting debate in the legal literature. *See generally* PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE, *supra* note 43, at 62–87.

63. *See* COHEN, INDUCTION & PROBABILITY, *supra* note 23, at 58–70.

human behavior, and denies only the possibility that one can objectively evaluate the truth of such probability statements.[64] The subjectivist view, similarly, does not deny the possibility that most people will agree in their assessments of the probabilities of most events.[65] But there is nothing in the theory that would require such consensus to form, nor that provides any basis for criticizing or refuting anyone who held a dissenting view of the probability of the same event.

C. Indifference Theory and Bayesianism

Two other accounts of the meaning of probability statements are also worth brief consideration. One is the classical or indifference theory of probability, which involves an application of the famous and controversial principle of insufficient reason.[66] Under this principle, differing outcomes are assigned equal degrees of probability if there is no available reason to judge one outcome as more likely than another.[67] Accordingly, when we assign each number on a die a one-sixth probability of turning up on any given roll, we do so because we have no information that would cause us to judge any of the six outcomes as more likely than the other.

The principle of insufficient reason is not justified either as deductively necessary nor inductively verifiable. Rather, it is a potential rule of decision theory that is logically and inductively permissible and that many have argued serves to minimize errors in appropriate circumstances—for example, in minimizing losses in games of chance.[68] While it is unlikely that the indifference theory is the best explanation of probabilistic statements concerning either potential health risks or litigation outcomes, it may have a significant—if frequently unacknowledged—role in justifying assertions of individualized risk derived from group-related frequentist probabilities. For example, suppose I have a class of one hundred students, fifty men and fifty women. Suppose I am told that there is a student waiting to see me in my office. What is the probability that the student is a man? The claim that fifty percent is the correct answer to that question can be justified as an application of the indifference principle. I did not say the student was "randomly selected" (which would justify the fifty percent assertion on a purely frequentist basis). Rather, I simply know nothing about the student other than that he or

64. *See id.*
65. *See id.*
66. *See, e.g.*, Kaye, *supra* note 43, at 3–4.
67. *Id.*
68. Ariel Porat et al., *Indeterminate Causation and Apportionment of Damages*, 23 OXFORD J. LEGAL STUD. 667, 690 (2003) (noting that the principle of insufficient reason is the "rational" assumption to make in predicting the results of a series of fair coin tosses, and arguing that the same assumption should be applied in cases of indeterminate causation in tort); *see also* D. H. Kaye, *The Error of Equal Error Rates*, 1 LAW, PROBABILITY & RISK 3 (2002) (arguing that the principle of insufficient reason should not be used to justify the preponderance of the evidence standard in civil cases).

she is in my class and has come to see me in my office. Justifying a fifty-percent assignment on the basis of the fact that I lack further individuating information, however, requires an appeal to the indifference principle.

Of course, it will be rare that we only know one characteristic about any real individual, and individuating characteristics could justify assigning different probabilities to that particular individual, even if we accept the indifference principle. If, in my experience, men are more likely to visit me during office hours than women, then I would be justified in assigning a probability of greater than fifty percent that the student in question was a man. How much greater, however, leads into a discussion of Bayes's Theorem and some important issues in behavioral theory.

Bayes's Theorem is a mathematical rule for combining a prior probability distribution for some individual property or event *A*, with a "likelihood function"—the general likelihood of *A* given the observance or observation of *B*.[69] Bayes's Theorem enables us to compute the posterior probability or specific likelihood of *A* in a given population after an observation of event or property *B*.[70] In my student visit hypothetical, when the distribution of men and women in the class is equal (fifty-percent prior probability that *A*, the student is male) then the likelihood function that male students are twice as likely to visit me as female students (the likelihood of *B*, given condition *A*) yields a posterior probability of 2/3 that the student who visited me was male (likelihood of *A*, given condition *B*). Suppose my class is seventy-five percent female and twenty-five percent male, but the male students are still twice as likely to visit as female students. Now Bayes's Theorem tells us that the posterior probability that the student visiting me is male, while more than twenty-five percent (the prior probability), is far less than sixty-six percent. In fact, it is forty percent.

While Bayes's Theorem provides a mathematical justification for combining information about frequencies (the likelihood function) with prior probabilities to arrive at a correct determination of the posterior probability of an individual event, it is agnostic with respect to the meaning of such probability judgments. Proponents of Bayes's Theorem argue it provides a normatively correct way of reasoning about all probabilities, whether derived from frequentist data, subjective belief, or the indifference principle.[71]

One of the interesting discoveries of behavioral theory is that human reasoning about probability rarely follows Bayes's Theorem.[72]

69. For a fuller discussion of Bayes's Theorem and its relation to classical theories of statistical inference, see Kaye, *supra* note 43.

70. *Id*.

71. *See, e.g.*, *id*. at 5.

72. *See* Ward Edwards, *Conservatism in Human Information Processing*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 359, 361–69 (Daniel Kahneman et al. eds., 1982) [hereinafter HEURISTICS & BIASES].

There are numerous studies in which individuals, after being given a prior probability distribution and a likelihood function, are asked to estimate the posterior probability, and generally fail to answer in accordance with Bayes's Theorem.[73]

Is it appropriate to say that such people are acting (or reasoning) irrationally? This is a subject of substantial debate among probability theorists.[74] While most would agree that Bayes's Theorem provides an uncontroversial and empirically verifiable way to reason about frequentist probabilities, the appropriateness of applying the same mode of reasoning to the probability of individual real world events remains an open question. Many have questioned whether the move that Bayes's Theorem authorizes from a prior probability distribution to the likelihood of the occurrence of a specific event is appropriate in contexts requiring individualized proof, such as a civil trial.[75] Some have gone even further and argued that trials and similar decision-making contexts appropriately utilize an entirely different method of probability assessment, sometimes referred to as "Baconian," the basic premises of which will be described shortly.[76]

Of even greater importance for this essay, however, is the recognition that Bayes's Theorem, while providing a method for combining prior probabilities with likelihood functions, provides no guidance for determining the correct values of either. In most of the experiments conducted by the behavioral economists, these values were simply stipulated, and the subjects were expected to accept and utilize them.[77] In the real world of policy debates, however, prior probability distributions, and especially likelihood functions, are frequently controversial and subject to uncertainty and debate. This is due not only to potential errors of measurement, but to the fact that every calculation of frequencies within a group must necessarily ignore individual characteristics of members of the group.[78] In theory, this additional individuating information can be incorporated into the Bayesian calculus through another likelihood function (i.e., the likelihood of A given the observation of new individuating characteristic B_1), but there may be little relevant data from which to compute such a likelihood function.[79] The question for policymakers is,

73. *Id*.

74. *See, e.g.*, *infra* Part I.E.

75. These concerns are perhaps best illustrated by the vexing "Paradox of the Gatecrasher," a hypothetical extensively analyzed and debated in PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE, *supra* note 43, at 29. It posits a rodeo in which 499 tickets are sold, but 1000 spectators are in the seats, authorizing the Bayesian inference that there is a .501 likelihood that each spectator is a gatecrasher and therefore liable for the ticket price in a civil lawsuit. *Id*.

76. *See infra* Part I.E.

77. *See*, *e.g.*, *infra* Parts II.A–B.

78. For example, in the Gatecrasher paradox, some have argued that additional individuating evidence might lead to a subjective probability of liability below fifty percent. PROBABILITY AND INFERENCE IN THE LAW OF EVIDENCE, *supra* note 43, at 30.

79. *See* L. Jonathan Cohen, *Bayesianism Versus Baconianism in the Evaluation of Medical Diagnoses*, 31 BRIT. J. FOR PHIL. SCI. 45, 52 (1980) [hereinafter Cohen, *Bayesianism Versus Baconianism*].

therefore, likely to be whether calculations of the likelihood of individual events based on estimates of prior probabilities and likelihood functions are more valid and reliable than direct estimates of the probabilities of the specific event in question after considering all known facts about it.

D. Propensity

Another concept of probability is reflected in the propensity theory, which views probability statements as descriptions of some presently existing attribute or characteristic of the thing being described.[80] Under this view, when we say that a die has a one-sixth likelihood of rolling six, we are making a statement about an inherent physical property of the die, not of the relative frequency of such rolls or our subjective belief or lack of knowledge about them. There is something tantalizingly appropriate about the propensity theory. It often seems best able to capture our intuitive sense of what it means for a future event to be probable. When we speak of a substance or activity as hazardous or risky, we do feel that we are somehow describing its propensity to cause harm. Even in describing the probable outcome of a lawsuit, the propensity theory has considerable appeal. If I state, for example, that a recently filed complaint is "likely" to be dismissed for failure to state a claim, then I seem to be, and feel myself to be, making an assessment about the objective characteristics of the complaint and its propensity to cause a decision maker to render a particular outcome.[81]

The fundamental difficulty with the propensity theory, however, is that while it may comport with our intuitive understanding of probability in many cases, it provides no way to assess those probabilities or to evaluate the accuracy of any such probability statement. As L. Jonathan Cohen notes, "The main weakness of a propensity analysis is that it does not intrinsically carry with it any distinctive type of guidance in regard to the actual evaluation of probabilities."[82] The indifference, frequentist, and subjectivist theories of probability each set forth circumstances under which particular probability assessments can be justified as accurate within the confines of the theory. The propensity theory, in contrast, simply asserts that there are properties of things called propensities that we can somehow perceive and measure, but gives no guidance as to how such judgments can be made or evaluated. Of course, this lack of direct empirical verifiability does not negate the sense that propensities seem to be what we are talking about in many of our discussions of probability. As we shall see, it is very hard to discuss a subject like "product risk"

80. *See, e.g.*, Kaye, *supra* note 43, at 4–5.

81. As a description of probability statements relating to uncertain past events, however, the propensity theory seems quite awkward. If we state that there is a fifty-percent probability that defendant committed the murder, then we do not mean he had equal propensities to do it and not to do it, but are more likely making a subjectivist statement about our degree of reasonable belief.

82. COHEN, INDUCTION & PROBABILITY, *supra* note 23, at 56.

without referring, implicitly or explicitly, to propensities. Indeed, with respect to some issues, frequentist and subjective probability assessments may be more appropriately viewed not as alternative probability concepts, but as indirect means of arriving at a valid assessment of propensities.

E. Weight and Baconian Probability Judgments

Consider the following two scenarios: You are teaching a first-year law school class on torts. At the end of the class, a student comes up and tells you that he was recently injured by falling into a pothole while skateboarding. He asks what his chances are of succeeding in a suit against the city. You tell him that you know of no case involving precisely those facts but, based on your general knowledge of tort law and municipal liability, your "off the cuff" judgment is that it is a tough case to win. If you had to bet, you would give odds of no better than 1/3.

In the second scenario, you are a practicing lawyer. A client comes to you with the same legal problem. You interview the client for an hour concerning the precise facts of the accident. You research all the pothole cases and all the skateboard cases you can find from the lower courts. You check the most recent appellate court rulings about municipal liability. You go out and inspect the pothole. At the end of your research, you conclude that it is a tough case to win. If you had to bet, you would give odds of no better than 1/3.

Most people, and certainly most lawyers, would be inclined to say that the second probability assessment is better than the first. It involves a more considered judgment based on more information. One would be hesitant to rely on the "off the cuff" legal impression of scenario one, whereas the judgment reached in scenario two is intended to form the basis for some course of action. The problem, of course, is that the probability assessment in both cases is exactly the same. In both cases we wind up assigning a probability of .33 to a successful outcome. There seems no way, using the language of Pascalian mathematical probability alone, that we can reflect the fact that the second judgment seems surer and more reliable than the first.

Some theorists have argued that in order to deal with this problem, a concept in addition to probability is required, a concept generally referred to as "weight."[83] When we speak of the weight of an argument we seem to be referring to the total amount of evidence supporting a probability judgment, even if that evidence is distributed relatively evenly for and against a particular outcome. In this view, weight and probability are independent of each other. Weight always increases as additional evidence is gathered, whereas probability may increase or decrease on

83. The concept of weight was first developed by John Maynard Keynes. J. M. KEYNES, A TREATISE ON PROBABILITY 71 (1921).

the basis of additional evidence. Moreover, unlike probability statements, statements about weight are not subject to the Pascalian probability calculus.[84] While it does seem possible to compare the weights of various probability judgments, and therefore to ordinally rank them, there is no theoretically satisfactory method whereby weights may be measured or assigned mathematical values.[85] Yet it is worth noting that much of what both lawyers and scientists purport to do, each in their respective field, is evaluate the weight of the evidence put forward to support various uncertain factual propositions.

The inability of Pascalian mathematical probability to incorporate the concept of weight has led some probability theorists to argue that it fails to capture some of our most important intuitive understandings about probability judgments.[86] They argue that in certain contexts, particularly those where questions of weight are paramount, an alternative non-Pascalian method of probability assessment is both normatively appropriate and the method actually used by most people and social institutions.[87] L. Jonathan Cohen, who has provided the most complete description of this alternative method of probability assessment, calls it "Baconian," after Francis Bacon, who developed one of the earliest theories of scientific induction based on causal laws.[88] Cohen stated the fundamental distinction between the two methods as follows: "Pascalian functions grade probabilification *on the assumption that* all relevant facts are specified in the evidence, while Baconian ones grade it *by the extent to which* all relevant facts are specified in the evidence."[89]

Baconian reasoning about probability proceeds from the "Method of Relevant Variables," in which the causal generalization implied by the probability judgment (which may be considered the hypothesis) is graded or ranked according to the "cumulatively increasing combinations" of relevant variables that support it.[90] For example, when William Safire made his June 2001 prediction that Gore was the frontrunner for the Democratic nomination,[91] it was probably based on a causal generaliza-

84. *See* COHEN, INDUCTION & PROBABILITY, *supra* note 23, at 40–42.

85. There is, however, an extensive experimental psychological literature that seeks to determine the way individuals assess the weight of various types of evidence. *See* Michael J. Saks, *Flying Blind in the Courtroom: Trying Cases Without Knowing What Works or Why*, 101 YALE L.J. 1177, 1188–91 (1992) and authorities cited therein at nn.50–60. While most of this work deals with "weight" in the standard lawyerly sense of persuasive value, rather than making the distinction between probability and weight cited above, it is not inconceivable that experiments could be designed to study the psychological assessment of "weight" in this more specialized sense. *See* Linda Stanley & Don Coursey, *Empirical Evidence on the Selection Hypothesis and the Decision to Litigate or Settle*, 19 J. LEGAL STUD. 145, 157 (1990) (giving the results of an experimental study indicating that increases in sample size of evidence relevant to the dispute led to higher settlement rates).

86. *See* Cohen, *Psychology of Prediction, supra* note 7, at 388–96.

87. *Id.*

88. *Id*. at 388.

89. *Id*. at 389.

90. COHEN, INDUCTION & PROBABILITY, *supra* note 23, at 145.

91. Safire, *supra* note 1.

tion something like "presidential nominees who lose close elections are renominated by their parties in subsequent elections." While this generalization may have some empirical support (Adlai Stevenson, Richard Nixon), it is clearly not very strong or reliable. It fails to take account of many other variables that are relevant to the choice of a presidential nominee. One important variable could not have been known by Safire at the time he wrote his article: the fact that Gore later announced he would not seek the 2004 nomination. Accordingly, a prediction that Gore would not be the nominee, based on the generalization that "presidential nominees who lose close elections, but then announce that they will not seek the nomination in the subsequent election are not renominated," accounts for more relevant variables and may therefore be considered more reliable,[92] in a Baconian sense, than Safire's original prediction.

The rules for reasoning about Baconian probabilities can be formalized, and they vary considerably from those relating to Pascalian probability. A basic feature of Pascalian probability is the complementational principle of negation, under which the probability of the occurrence of an event $p(e)$ and the probability of its nonoccurrence, $p(not\ e)$ will always equal 1. Accordingly, if $p(e) = n$, then $p(not\ e) = 1 - n$.[93] The same relation does not hold with respect to Baconian probabilities.[94] If we had only a small amount of evidence relating to the 2004 Democratic nomination, then the Baconian probability that Gore would be the nominee and that he would not be the nominee could both be considered low.[95] Moreover, the multiplication principle of Pascalian probability, on which Bayes's Theorem is based, does not apply to Baconian probabilities, and therefore neither does Bayes's Theorem.[96]

92. In his writing about Baconian probability, Cohen used the term "inductive reliability," or "inductive support," rather than weight. COHEN, INDUCTION & PROBABILITY, *supra* note 23, at 159–60; Cohen, *Psychology of Prediction*, *supra* note 7, at 389. Alex Stein uses the term "resiliency" of a probability assessment to describe its susceptibility to revision "in the event of change in its underlying informational base." Alex Stein, *An Essay on Uncertainty and Fact-Finding in Civil Litigation, With Special Reference to Contract Cases*, 48 U. TORONTO L.J. 299, 306 (1998). Probability judgments that are better able to withstand such changes are said to be "more resilient, and correspondingly, more reliable." *Id*.

93. *See* IAN HACKING, AN INTRODUCTION TO PROBABILITY AND INDUCTIVE LOGIC 166 (2001) [hereinafter HACKING, INDUCTIVE LOGIC].

94. Accordingly, William Safire's June 2001 predictions about potential presidential candidates are not subject to the criticism made at the beginning of this essay—that they violate the Additivity Rule—if viewed as Baconian rather than Pascalian probability statements.

95. *See* COHEN, INDUCTION & PROBABILITY, *supra* note 23, at 104–05, 157–59.

96. The basic multiplication rule states that the probability of two independent uncertain events is the product of their probabilities. For example, the probability of rolling six on a single die twice is 1/6 *x* 1/6, or 1/36. Bayes's Theorem uses this same principle to combine prior probabilities with conditional probabilities. HACKING, INDUCTIVE LOGIC, *supra* note 93, at 42, 69–71. There is no equivalent of the multiplication rule for Baconian probability, and "[i]n Baconian reasoning prior probabilities always set a floor to posterior ones, but never a ceiling." Cohen, *Psychology of Prediction*, *supra* note 7, at 392.

The proponents of Baconian probability do not view it as a refutation or potential replacement for Pascalian probability.[97] Rather, they see it as an alternative and complementary approach to probability assessment, with different strengths and weaknesses, which may be more appropriate to use in certain contexts.[98] One of their most striking claims is that Baconian probability more accurately describes the way most people (and important social institutions, such as courts) reason about the probability of individual events.[99] For example, the basic legal rule that each element of the claim in a civil trial must be proved by a preponderance of the evidence has always been difficult to justify on Pascalian principles of probability, but seems quite consistent with Baconian reasoning.[100] Similar claims about the intuitive and normative plausibility of Baconian reasoning about probability underlie Cohen's critique of the Kahneman and Tversky studies, which will be discussed in the following section of this essay.[101]

F. Risk and Uncertainty

Another longstanding issue in probability theory is the purported distinction between risk and uncertainty.[102] The distinction, first put forth by Frank Knight, is between known and unknown probabilities.[103] If I know a jar contains one hundred marbles, fifty black and fifty white, then I know that the probability of selecting a white one at random is fifty percent. If I know that a jar contains one hundred marbles, and that they are white and black marbles in some unstated proportion, then I may, by applying the indifference principle, still conclude that the chance

97. *See* Cohen, *Bayesianism Versus Baconianism, supra* note 79, at 58–59.

98. *Id*.

99. *See* Cohen, *Psychology of Prediction, supra* note 7, at 393–96.

100. If there are two elements to the claim, say negligence and causation, and each is proved with a probability of .6, then the plaintiff should win. Yet under the multiplication rule of Pascalian probability, if negligence and causation are independent variables, then the probability of liability in such a case is only .36. Cohen points out that under Baconian principles, the "probability of a conjunction does not have to be less than the probability of either of its conjuncts." *Id*. at 391. Thus, the probability that "defendant was negligent and its negligence caused the accident" can be equal to the probability that "defendant was negligent" and "its negligence caused the accident." Cohen argues that this is the reasoning utilized by most judges and jurors and leads him to conclude that "the Anglo-American legal system endorses a Baconian, rather than a Pascalian, framework of reasoning in this context." *Id*. at 391, 395.

Alex Stein points out that this rule applies only to "elemental facts"—the lawsuit's constitutive elements. Alex Stein, *Of Two Wrongs That Make a Right: Two Paradoxes of the Evidence Law and Their Combined Economic Justification*, 79 TEX. L. REV. 1199, 1205 (2001). He distinguishes them from the intermediary facts that are used to establish those elements and that follow Pascalian principles in doing so, thereby casting some doubt on Cohen's broad claim for the prevalence of Baconian reasoning in Anglo-American law. *Id*. Stein proposes that the "Conjunction Paradox" is justified as a way of preventing potential inaccuracies in adjudication that would otherwise result from courts' attention to ex post rather than ex ante probabilities of breach. *Id*. at 1202–05.

101. *See infra* Part II.C.

102. *See* FRANK H. KNIGHT, RISK, UNCERTAINTY AND PROFIT 214–15 (1921).

103. *Id*.

of selecting a white one at random is fifty percent. The first scenario is an example of risk, the second an example of uncertainty. While the distinction can frequently be useful, particularly in mathematical contexts, in the real world the difference between risk and uncertainty is almost invariably a matter of degree. Except for the gambling tables at Las Vegas, there are few stochastic processes where risks can be perfectly ascertained, but there are similarly few real-world events where some frequency estimates and base rate statistics cannot be compiled. Accordingly, the distinction will mostly be ignored in this essay.

II. PROBABILITY AND THE AVAILABILITY HEURISTIC

The availability heuristic is one of the central concepts of behavioral theory. It has been articulated in numerous ways in recent legal scholarship, with most, but not all of those formulations relating it to the way individuals make probability judgments.[104] Kahneman and Tversky, whose work is almost invariably cited as the source of the concept and for empirical evidence of its use, describe it as follows: "A person is said to employ the availability heuristic whenever he estimates frequency or probability by the ease with which instances or associations could be brought to mind."[105]

This section looks closely at Kahneman's and Tversky's original studies from the perspective of probability theory. It will show, as the previously quoted statement indicates, that Kahneman's and Tversky's experimental work on the availability heuristic related solely to determinations of frequency and frequentist probability, and that Kahneman and Tversky clearly recognized this. In fact, in the fifth and final part of their seminal paper, they addressed the estimation of probabilities of unique events, and suggested (but provided no evidence) that the availability heuristic operates there in a somewhat different way than it does with frequentist probabilities.[106] This section will then consider Cohen's important but much less well-known critique of Kahneman's and Tversky's studies.[107]

104. For example, Sunstein speaks of the "availability heuristic," in accordance with which people assess the probability of an event by seeing whether relevant examples are cognitively "available." Sunstein, *supra* note 53, at 1125; *see also* Mark Geistfeld, *Placing a Price on Pain and Suffering: A Method for Helping Juries Determine Tort Damages for Nonmonetary Injuries*, 83 CAL. L. REV. 773, 836 (1995) (noting studies showing that individuals tend to disregard objective probabilities when risks are vivid or salient).

105. Amos Tversky & Daniel Kahneman, *Availability: A Heuristic for Judging Frequency and Probability* [hereinafter Tversky & Kahneman, *Availability*], *in* HEURISTICS & BIASES, *supra* note 72, at 163, 164.

106. *Id.* at 175–78.

107. *See infra* Part II.C.

A. Kahneman's and Tversky's Classic Studies of Availability

The studies which formed the basis for Kahneman's and Tversky's famous article on availability involved a series of word, math, and memory puzzles.[108] In the first study, participants were given seven seconds to estimate the number of words they could construct in two minutes from a matrix of nine letters.[109] In the second study, participants were given seven seconds to estimate how many instances of a certain category (for example, flowers or Russian novelists) they could recall in two minutes.[110] In both cases, participants' estimates of their likely performance on these tasks closely correlated with their subsequent actual performances.[111] Kahneman and Tversky concluded that these "studies show that people can assess availability quickly and accurately."[112] In other words, people can accurately assess their ability to retrieve available instances of an event or occurrence, real or imagined, without actually doing so.

In the third study, participants were asked to estimate the objectively verifiable frequency of an occurrence, instead of their own ability to retrieve instances of the occurrence.[113] Participants were told that there had been a study of "[t]he frequency of appearance of letters in the English language" and: "A typical text was selected, and the relative frequency with which various letters of the alphabet appeared in the first and third positions in words was recorded. Words of less than three letters were excluded from the count."[114]

The participants were then asked to "consider" various letters (such as "R") and asked whether they thought the letter was more likely to appear in the first or third position.[115] They were also asked to estimate the ratio of these two values.[116] Extensive word count studies have established that the majority of consonants appear more frequently in the first than in the third position in most English texts.[117] Eight consonants ("X," "Z," "D," "K," "L," "N," "R," "V") are exceptions to this rule, and appear more frequently in the third position.[118] Kahneman and Tversky eliminated "X" and "Z" on the grounds that they were "relatively rare," and "D" because it was "more frequent in the third position only in three letter words."[119]

108. Tversky & Kahneman, *Availability*, *supra* note 105, at 165–68.
109. *Id.* at 165.
110. *Id*.
111. *See id*. at 165–66.
112. *Id*. at 166.
113. *Id*. at 166–68.
114. *Id*. at 167.
115. *Id*.
116. *Id*.
117. *Id*.
118. *Id*.
119. *Id*. It may also be the case that more words with these letters in the third position were available to most participants than words with these letters in the first position. Compare "mad", "sad", "had", "box", "fuzz", and puzzle; to "dog", "xylophone", and "zebra".

The other five consonants, "K," "L," "N," "R," "V", were the ones which the relative frequency participants in the study were asked to judge.[120] Of 152 participants, 105 incorrectly judged the first position to be more likely for a majority of these letters.[121] Forty-seven judged the third position more likely for a majority of the letters.[122] Moreover, each of the five letters was judged by a majority of the participants to be more frequent in the first position.[123]

This study, which is often cited in the behavioral economics literature, does indeed provide strong, if indirect, evidence that people use availability to judge relative frequency. The hypothesis that most people can more easily recall instances of words beginning with "K," "L," "N," "R," and "V" than they can recall words with those letters in the third position seems the best explanation for the fact that most participants in the study judged the relative frequency of the positions of these letters incorrectly. Yet two other important facts about this study must also be noted. First, the study involved the five English consonants where use of the availability heuristic was most likely to lead to incorrect frequency judgments. If the participants had been asked to assess the relative frequencies of "B," "M," or "S" in the first and third positions, then they would likely have used the availability heuristic to conclude, correctly, that the first position was the more frequent. I suspect that if asked about the letter "X," the greater availability of words with that letter in the third position would also have led most participants to the correct result. This in no way impugns the validity of Kahneman's and Tversky's study, which was designed to show that the availability heuristic is in fact used in making frequency judgments and not to assess the accuracy or usefulness of that heuristic. In light of the disparagement of the availability heuristic frequently seen in the law and behavioral theory literature, however, it is worth noting that if one uses the availability heuristic to judge the relative frequency of all English letters in the first and third positions, then one is likely to be correct most of the time.

The second point about the Kahneman and Tversky study is that it dealt with judgments concerning empirically verifiable facts, so that the accuracy of those judgments could be relatively easily assessed. The study was based on extensive word count tables, which had been published for use in psychology experiments.[124] These provide an objective and highly reliable indication of the actual frequency of the occurrences. As Kahneman and Tversky were well aware, the effect and usefulness of availability are more difficult to assess when dealing with probabilities of individual events.[125]

120. *Id*.
121. *Id*.
122. *Id*.
123. *Id*.
124. *Id*.
125. *Id*. at 175–76.

The results of the study of the relative frequency of letter position were essentially duplicated in three other studies involving geometric and mathematical puzzles.[126] In one study, participants were shown two different two dimensional "structures" and were asked which provided more paths to connect the top of the structure to the bottom.[127] In another, participants were asked whether, out of a group of ten people, it was possible to form more different committees consisting of two people or of eight people.[128] In the third study, two different groups were asked to estimate the value of either (*8 x 7 x 6 x 5 x 4 x 3 x 2 x 1*) or (*1 x 2 x 3 x 4 x 5 x 6 x 7 x 8*).[129] In the first study, although there were in fact an equal number of paths through each structure, most people "saw" more paths through the first because its shape made those paths more available.[130] Similarly, in the second study, an equal number of different committees may be composed of either two or eight out of ten potential members, because each unique committee of two also defines a unique committee of eight.[131] Nonetheless, a majority of the participants estimated that a greater number of committees could be composed of fewer members.[132] Finally, the estimation of the value of the multiplication when written in descending number order (median estimate 2215), was much higher than when it was written in ascending order (512), yet both were far lower than the actual value (40,320).[133] Kahneman and Tversky explained these results by suggesting that most people estimate final results from extrapolating from the first few available instances.[134] Because it is easier initially to "see" the different two-person committees than those with eight members, it is assumed that two-person committees will predominate. Similarly, because the first few multiplications yield a smaller number when arranged in ascending order than descending order, the extrapolation from those first few multiplications also yields a smaller estimation of the total result.[135]

A study of fame and recall provided similar results.[136] Participants were read a list of names of known personalities.[137] Some of the lists consisted of nineteen very famous men (such as Richard Nixon), and twenty

126. *Id*. at 168–71.
127. *Id*. at 168.
128. *Id*. at 169.
129. *Id*. at 170–71.
130. *Id*. at 168.
131. *Id*. at 169.
132. *Id*.
133. *Id*. at 170–71.
134. *Id*. at 171.
135. Again, it is worth noting that these questions were carefully set up so that the presumed use of availability in judging frequency would yield incorrect results. If the question had been whether more committees of three or nine members can be formed from ten people, then the greater availability of different three member committees would have been the correct answer.
136. Tversky & Kahneman, *Availability, supra* note 105, at 175.
137. *Id*.

somewhat less famous women (such as Lana Turner).[138] The others contained nineteen very famous women (such as Elizabeth Taylor), and twenty somewhat less famous men (such as William Fulbright).[139] Of the ninety-nine participants asked to judge the relative "frequency of men and women in the list, 80 erroneously judged the class consisting of the more famous names to be more frequent."[140]

Taken together, these studies do indeed provide powerful support for the idea that the ease with which an event or occurrence is imagined or recalled can have a strong effect on the estimation of its frequency. Just how strong that effect will be, particularly when other methods for estimating frequency may yield different conclusions, was examined in another study described by Kahneman and Tversky, which they captioned "availability vs. representativeness."[141]

In this study, participants were again presented with a "structure" which looked this:

X X O X X X
X X X X O X
X O X X X X
X X X O X X
X X X X X O
O X X X X X

Participants were then asked to estimate the relative frequency of paths that could be drawn from the top to bottom of the structure.[142] They were asked to estimate the percentage of paths that would go through 6 Xs and no Os, those that would go through 5 Xs and 1 O, down to no Xs and 6 Os.[143] Again, there is a right answer to this question and it is that the highest number of possible paths, forty percent of the total, go through 5 Xs and 1 O.[144] More than 2/3 of the participants, however, incorrectly estimated that the most paths would go through 6 Xs and no Os.[145] The apparent explanation for these mistakes is, again, that paths involving no Os are easier to "see" when looking at the structure, and are therefore presumed by most participants to be the most frequent.[146]

Kahneman and Tversky then tested a different version of the same mathematical problem.[147] Participants were told: "Six players participate in a card game. On each round of the game, each player receives a single card drawn blindly from a well-shuffled deck. In the deck, 5/6 of the

138. *Id*.
139. *Id*.
140. *Id*. at 175.
141. *Id*. at 171.
142. *Id*.
143. *Id*.
144. *Id*. at 172.
145. *Id*.
146. *See id.*
147. *Id*. at 173.

cards are marked X and the remaining 1/6 are marked O."[148] The participants were then asked to estimate, in repeated rounds of the game, the relative frequency of rounds in which all six players received X cards, 5 received X and 1 O card, etc.[149] When asked this question, seventy-one of the eighty-two participants correctly identified 5 Xs and 1 O as the most frequent result.[150]

Kahneman and Tversky hypothesized that the latter problem elicited a mode of evaluation from the participants different from availability—the representativeness heuristic.[151] As they stated:

> In the path problem, individual instances were emphasized by the display, and the population proportion (i.e., the proportion of Xs in each row) was not made explicit. In the card problem, on the other hand, the population proportion is explicitly stated and no mention is made of individual instances. Consequently, we hypothesize that the outcomes in the card problem will be evaluated by the degree to which they are representative of the composition of the deck rather than by the availability of individual instances.[152]

The lesson of the "availability vs. representativeness" study, the results of which have been borne out by much subsequent research, is that the availability heuristic is not some hardwired bias that necessarily causes people to think irrationally about frequency and probability. Instead, it is a tentative decision-making tool that can be relatively easily elicited or suppressed, depending on how information is presented and questions are posed. In the Kahneman and Tversky study, the right answer was generated by the representativeness heuristic, which could be elicited in most people by emphasizing the general characteristics of the group or population in question rather than individual instances or members of the group.

In this view of things, behavioral theory poses as much of a challenge to the responsibility of media organizations as it does to the assumptions of rational choice theory. By overemphasizing individual events at the expense of broader analyses of trends and general social phenomena, media coverage can invoke in people unreasonable fears due to the greater availability of vivid, but relatively rare, events and images. Some work in behavioral theory and law has indeed taken this path, discussing the potentially deleterious effects of "availability cascades" or "overwhelmingly informative" conditions on individuals' perceptions of risks.[153] Although this concern about the potentially distort-

148. *Id*.
149. *Id*.
150. *Id*. at 174.
151. *Id*. at 173.
152. *Id*.
153. *See, e.g.*, Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683, 683 (1999); Shelly E. Taylor, *The Availability Bias in Social Perception and Interaction*, *in* HEURISTICS & BIASES, *supra* note 72, at 190, 195.

ing effect of modern media coverage on public policy debates may have validity in some cases, it hides the more basic problem.

All the questions posed by the Kahneman and Tversky studies had determinate and determinable correct answers. It was the knowledge of those correct answers that led us to say, in many of the studies, the availability heuristic caused participants to overestimate the frequency of some occurrences and to underestimate the frequency of others. Similarly, in the availability vs. representativeness study, it is that same knowledge of the right answer that enables us to say that representativeness, not availability, is the heuristic that should be invoked. In the absence of any generally accepted correct answer to questions of probability, however, it is impossible to say whether any particular form of media coverage or presentation of information is more likely to lead to correct or incorrect public perception of these issues.

For example, does the breaking off of a piece of the Antarctic ice shelf the size of Rhode Island provide a vivid illustration of the perils of global warning (the availability heuristic used to accurately convey information about a broader and more complex set of events),[154] or is it a misleading individual event that fails to provide an accurate picture of many different and contradictory changes taking place in the Antarctic climate (a misuse of the availability heuristic)?[155] Once we recognize, as Kahneman and Tversky clearly do, that it is not the use of the availability heuristic that is the problem, but its use in situations where other decision-making techniques will yield more accurate results, we are left with the question of how to use the availability heuristic in situations where we have no objective way to determine the correct answer. Judging the probability of an individual event is one such situation.

B. Kahneman and Tversky on Subjective Probability Judgments

In the very same paper in which they present the evidence for the availability heuristic, Kahneman and Tversky present some suggestions to deal with precisely this problem.[156] Recognizing that objectively correct answers are not available in "many real-life situations where probabilities are judged," Kahneman and Tversky assert that "[n]evertheless, the availability heuristic may be applied to evaluate the likelihood of such events."[157] In this portion of their paper, however, they do not provide scientific experimental data, but rather offer to "discuss the role of

154. *See* Andrew C. Revkin, *Large Ice Shelf in Antarctica Disintegrates at Great Speed*, N.Y. TIMES, Mar. 20, 2002, at A13.

155. *See* Kenneth Chang, *The Melting (Freezing) of Antarctica; Deciphering Contradictory Climate Patterns Is Largely a Matter of Ice,* N.Y. TIMES, Apr. 2, 2002, at F1 (discussing possible interpretations of melting Antarctic ice).

156. Tversky & Kanheman, *Availability, supra* note 105, at 175–78.

157. *Id*. at 175–76.

availability in such judgments, speculate about expected sources of bias, and sketch some directions that further inquiry might follow."[158]

They begin with a heuristic quite familiar to lawyers and legal academics: the hypothetical. They envision an "imaginary clinical situation" in which a psychiatrist hears a patient complain that he is tired of life, and wonders whether the patient is likely to commit suicide.[159] They speculate that the clinician will search his or her memory for prior relevant cases, and they point out that a number of different criteria might be used for the memory search.[160] They ask:

> In scanning his past experience does the clinician recall patients who resemble the present case, patients who attempted suicide, or patients who resemble the present case *and* attempted suicide? From an actuarial point of view, of course, the relevant class is that of patients who are similar, in some respects, to the present case, and the relevant statistic is the frequency of attempted suicide in this class.[161]

Yet Kahneman and Tversky worried that the memory search "may follow other rules."[162] Because clients who attempt suicide are likely to be more memorable and therefore more available, the clinician may try to judge how similar this patient is to other patients who attempted suicide, or may consider how frequently those suicidal patients also suffered from depression.[163] Both these memory searches potentially ignore the relevant fact that most depressed patients do not attempt suicide.[164] Accordingly, Kahneman and Tversky considered such judgments flawed, even though they involve subjective and not frequentist concepts of probability.[165]

What Kahneman and Tversky are doing here is applying a nonmathematical version of Bayesianism to subjective probability judgments.[166] The reasoning of the clinician is flawed because it fails to take into account the base rate, or prior probability, which in this case is the frequency with which depressed people attempt suicide. It is only after consideration of that frequentist base rate that one should consider the subjective probability that any individual patient with particular characteristics is likely to attempt suicide.[167] In a real Bayesian calculation, the

158. *Id*. at 176.
159. *Id*.
160. *Id*. at 176–77.
161. *Id*. at 176.
162. *Id*.
163. *Id*.
164. *Id*. at 176–77.
165. *Id*.
166. Recall that proponents of Bayesianism believe that it is the appropriate method for arriving at rational judgments about subjective as well as frequentist probabilities. *See supra* note 71 and accompanying text; *see also* HACKING, INDUCTIVE LOGIC, *supra* note 93, at 163–69 (applying Bayesian principles).
167. In theory, Bayes's Theorem can be used to combine a subjective prior probability with a subjective likelihood function, but Kahneman's and Tversky's reference to "the frequency of attempted

prior probability would be expressed mathematically, and would then be modified by a likelihood function—the likelihood of *A* (the patient has these characteristics) given *B* (the patient is suicidal). Here, however, the concept of a likelihood function is problematic.

The hypothetical patient is (by hypothesis) an actual clinical subject whose characteristics can be analyzed at any level of specificity the clinician chooses. If the clinician decides nothing more can be known about the patient than that he is depressed, then the best estimate of the subjective probability of suicide is the base rate or frequentist probability at which depressed people generally attempt suicide. I doubt, however, that patients, or most clinicians, would be satisfied with a judgment made on this basis. They would expect a more individualized judgment based on the specific characteristics of this individual's psychological state and the clinician's prior experience. The more individualized such a judgment gets, however, the less it resembles a Bayesian likelihood function and the more it resembles the (supposedly) flawed method of comparing the patient's condition to other suicidal patients the clinician has known.[168]

Alternatively, Kahneman's and Tversky's concern about experts potentially ignoring base rates may be based not on a strong endorsement of Bayesianism, but simply on a cautious attitude toward the use of any heuristics by experts when serious probability judgments must be made. If so, then all they can ultimately say about individual probability judgments is that they should be made after a careful consideration of all relevant facts, including base rates, to whatever degree they are knowable. The danger of availability, representativeness, and all other heuristics is that they truncate the reasoning process, and therefore lead to the consideration of certain facts to the exclusion of others.

Kahneman's and Tversky's last comment on the potential flaws of individual probability judgments seems also to reflect this concern with oversimplification. They state:

> Some events are perceived as so unique that past history does not seem relevant to the evaluation of their likelihood. In thinking of such events we often construct *scenarios*, i.e., stories that lead from the present situation to the target event. The plausibility of the

suicide in this class" indicates their view that the best source of a prior probability in this hypothetical would be a statistical frequency. Tversky & Kahneman, *Availability*, *supra* note 105, at 176.

168. A Bayesian likelihood ratio involves a determination of *Pr(E/H)*, the conditional probability of observing evidence *E* under condition *H* (hypothesis is true). In the Kahneman and Tversky hypothetical, it would require determination of the likelihood of observing *E* (patient's individual symptoms) under condition *H* (patient is suicidal). This could then be combined with the base rate to determine the posterior probability that patients with such symptoms are suicidal. It is unlikely, however, in a clinical context involving analysis of the symptoms of an individual patient, that the psychiatrist's determination of *Pr(E/H)*—the likelihood that one would observe such symptoms given that the patient is suicidal—is any easier to determine or more valid than a direct assessment of *Pr(H/E)*—the likelihood that patient is suicidal given the observed symptoms. *See* Cohen, *Bayesianism Versus Baconianism*, *supra* note 79, at 52–53.

> scenarios that come to mind, or the difficulty of producing them, then serve as a clue to the likelihood of the event. If no reasonable scenario comes to mind, the event is deemed impossible or highly unlikely. If many scenarios come to mind, or if the one scenario that is constructed is particularly compelling, the event in question appears probable.[169]

The danger with this is that simple scenarios with few relevant factors are easier to call to mind than complex scenarios involving many interacting processes.[170] For this reason, simple extrapolations from current data often lead to inaccurate predictions, while computer simulations of complex systems often lead to counterintuitive results.[171]

Kahneman's and Tversky's warning about the dangers of subjective probability judgments based on availability and other heuristics makes sense, if understood as a warning against making snap or incomplete judgments, against failing to consider base rates and other relevant—if less vivid—facts, and against failing to consider complex as well as simple scenarios to the best of our ability. Yet they do not give us any real programmatic guidance in making such judgments.

Taking a broad, multifaceted approach to assessing the probability of single events might cause us to revise our preliminary probability judgments, but then again it might not. The hypothetical clinician might decide not to alter his or her initial assessments of the probability the patient would attempt suicide after carefully assessing the patient's behavior in light of the base rate for attempted suicides, his clinical experience with depressed patients who both did and did not attempt suicide, and possible scenarios, simple and complex, that could lead this patient to attempt suicide. What would necessarily have changed is the weight, in a Keynesian sense, that the clinician, and any objective observer, would assign to the clinician's probability assessment.

Accordingly, Kahneman's and Tversky's discussion of the effect of availability on subjective probability judgments may be viewed as a perfectly reasonable warning that heuristics (which are, after all, short cuts in reasoning), if used to the exclusion of other available methods of probability assessment, will lead to judgments that will be less reliable and entitled to less weight than those made after a full consideration of all relevant data.

C. Cohen's Critique of Kahneman's and Tversky's Studies

One of the most interesting critiques of Kahneman's and Tversky's studies of availability and probabilistic decision making was made by L.

169. Tversky & Kahneman, *Availability, supra* note 105, at 177.
170. *Id*.
171. *Id*.

Jonathan Cohen.[172] Cohen questioned whether human "irrationality" can ever be experimentally demonstrated, pointing out that human reasoning itself is the source of our concepts of "rationality" and that therefore "ordinary human reasoning . . . cannot be held to be faultily programmed: It sets its own standards."[173] Cohen distinguished between the claim that people make various kinds of mistakes in reasoning, which he viewed as fundamentally no different or more controversial than the claim that people make various kinds of mistakes in math, and the much more dubious claim that ordinary people's ability to reason, particularly about probabilities, is fundamentally and systematically flawed.[174]

Cohen viewed Kahneman's and Tversky's experiments on the availability heuristic as evidence of mistakes rather than fundamental defects in reasoning.[175] He distinguished between the "wild assumption" that "frequency can safely be taken to equal availability" and the more reasonable assumption that frequency of a characteristic in a population may be judged by the frequency of the characteristic in the available population.[176] The latter assumption may be correct or not, depending on how representative the available population is of the total population.[177] Even when it is wrong, however, Cohen argued that it is an error more akin to an optical illusion, in which our ordinary capacity for judging frequency has been "obstructed by factors like the recency or emotional salience of the existing evidential input, by the existence of competing claims for computing time, or by a preference for least effort."[178] But such mistakes no more prove that people are irrational in judging frequency than optical illusions prove that people are irrational in judging length or height. The proof of both is that once the error is explained to them, people can use this additional knowledge to make improved judgments.

With respect to some of the Kahneman and Tversky experiments involving probability judgments, however, Cohen offers a far more thoroughgoing critique.[179] He believes that, in judging irrational their subjects' failure to apply Pascalian and Bayesian principles in reasoning about probabilities, Kahneman and Tversky are applying an incorrect and unproven normative theory.[180] He considers at some length an ex-

172. *See* L. Jonathan Cohen, *Can Human Irrationality Be Experimentally Demonstrated?*, 4 BEHAV. & BRAIN SCI. 317, 317, 326 (1981) [hereinafter Cohen, *Human Irrationality*] (criticizing assertions that research suggests "bleak implications for human irrationality," and labeling Tversky and Kahneman "untutored in statistical theory"); Cohen, *Psychology of Prediction*, *supra* note 7; L. Jonathan Cohen, *Whose Is the Fallacy? A Rejoinder to Daniel Kahneman and Amos Tversky*, 8 COGNITION 89, 89–92 (1980) [hereinafter Cohen, *Whose Is the Fallacy?*].

173. Cohen, *Human Irrationality*, *supra* note 172, at 317.

174. *Id.* at 325.

175. *Id*.

176. *Id*.

177. *Id*.

178. *Id*.

179. *Id*. at 328.

180. *Id*.

periment in which subjects were told that "in a certain town blue and green cabs operate in a ratio of 85 to 15," and that a witness identified the cab involved in the accident as green, under light conditions where "he can distinguish blue cabs from green ones in 80% of cases."[181] When subjects were asked the probability that the cab involved in the accident was blue, the median estimated probability was .2.[182] Tversky and Kahneman viewed this a systematic error, a failure to consider the prior probability that eighty-five percent of the cabs in town were green, which lead to an error in computing probabilities in accordance with Bayesian principles, under which the correct probability is 17/29, or fifty-nine percent, that the cab was blue.[183]

Cohen challenged this conclusion, arguing that the subjects of the experiments were not incorrectly applying Bayesian principles of probability, but correctly ignoring irrelevant data.[184] He pointed out that the "prior probability" of eighty-five percent blue cabs is a frequency statement, relevant to the long run accuracy of a series of cab-color identification problems, but not necessarily to the correctness of an individual identification of a cab as green.[185] As Cohen explained:

> If the jurors know that only 20% of the witness's statements about cab colours are false, they rightly estimate the probability at issue as 1/5, without any transgression of Bayes's law. The fact that cab colours actually vary according to an 85/15 ratio is strictly irrelevant to this estimate, because it neither raises nor lowers the probability of a specific cab-colour identification being correct on the condition that it is an identification by a witness. A probability that holds uniformly for each of a class of events because it is based on causal properties, such as the physiology of vision, cannot be altered by facts, such as chance distributions, that have no causal efficacy in the individual events.[186]

Cohen offers, as another illustrative hypothetical, a situation in which you, the reader, have symptoms consistent with either disease *A* or *B*.[187] A medical diagnostic test that is correct eighty percent of the time indicates that you have disease *B*, but disease *A* is nineteen times more common in the general population than *B*.[188] Both diseases are fatal if untreated, and the treatments for the two may not be combined.[189] Is it rational to treat you for disease *A* or *B*? Cohen is quite sure that you

181. *Id*.
182. *Id*.
183. *Id*.
184. Cohen insists that his critique here does not rely on his Baconian theory of probability, but on an incorrect application of Bayes's Rule. *Id*. at 328–29.
185. *Id*.
186. *Id*.
187. *Id*. at 329.
188. *Id*.
189. *Id*.

would, and should, opt to be treated for *B*.[190] He views the frequency of *A* and *B* in the general population as irrelevant to the question at hand, which is "a propensity-type probability" involving "success in his own particular case."[191] Cohen argues that in the absence of individuating causal links between the frequency data and the individual patient, "[w]e have to suppose equal predispositions here, unless told that the probability of *A* is greater (or less) than that of *B* among people who share all your relevant characteristics, such as age, medical history, blood group, and so on."[192]

Most of the commentators on this hypothetical, however, did not reach the same conclusion. Daniel Kahneman dismissed Cohen's rejection of frequency data as "irrational," believing that Cohen had conceded that "the physician, who uses long run frequencies, will in the long run make more correct diagnoses."[193] Simon Blackburn, a philosopher, thought the main problem was in "the metaphysical notion of a propensity, thought of as a particular real, but gradable feature of individual trials on a chance setup."[194] David H. Krantz, a psychologist and statistician, pointed out that the reliability of the test result could be viewed as a statement about frequency, as well.[195]

It became reasonably clear from these comments, and Cohen's reply, that the dispute is really about when a prior frequency distribution is relevant to an individual probability judgment. Even Cohen is willing to agree that frequencies are relevant when they describe the incidence of the disease "among people who share all your relevant characteristics, such as age, medical history, [and] blood group," and believes that everyone would agree that frequencies are irrelevant when clearly not causally related to the individual probability at issue.[196] The question is how to determine the relevance of prior frequency determinations when our knowledge of the causation involved is weak or incomplete. It is here that disagreements and even paradoxes emerge.

Consider the author of a book asked whether it is likely to contain any errors. If she considers each and every statement in the book individually, she believes each one to be correct. When she thinks about the book as a whole, however, she decides to include in the preface an apology for all the errors a reader may find in it. Does she believe the book

190. *Id*.
191. *Id*.
192. *Id*.
193. Daniel Kahneman, *Who Shall Be the Arbiter of Our Intuitions?*, 4 BEHAV. & BRAIN SCI. 339, 340 (1981).
194. Simon Blackburn, *Rational Animal?*, 4 BEHAV. & BRAIN SCI. 331, 331 (1981).
195. David H. Krantz, *Improvements in Human Reasoning and an Error in L. J. Cohen's*, 4 BEHAV. & BRAIN SCI. 340, 341 (1981).
196. L. Jonathan Cohen, *Are There Any A Priori Constraints on the Study of Rationality?*, 4 BEHAV. & BRAIN SCI. 359, 365–66 (1981). His example is bilharzia, "one of the commonest diseases in the world," but one which would be "absurd" to consider in "diagnosing a patient who never wades in fresh water." *Id*.

to be error-free or not? This "paradox of the preface" has no generally accepted solution, but clearly the paradox arises from the different degrees of belief under uncertainty that the author has when she considers each statement individually and in the aggregate.[197] The frequentist probability judgment that "most books have some errors" seems more relevant to determining whether all statements in the book are correct, which itself calls for a frequency judgment, than it does in determining the probability that each particular statement in the book is correct, which involves individual subjective probabilities. In the following sections, we will see that many real-life policy disputes create similar dilemmas.[198]

Although Cohen's critique of the Kahneman and Tversky experiments is intriguing and frequently illuminating, in the end it proves both too little and too much. Describing heuristics such as availability as involving mere errors in our capacity to judge frequencies instead of as fundamental defects in human reasoning is surely a justifiable and useful corrective to overblown claims about human irrationality. But even ordinary errors, particularly common and predictable ones, may have important consequences for policymakers. For example, it may be just a common mathematical error that many do not understand the difference between an average and a median, but such predictable misunderstandings seem to have been helpful to the Bush administration in building public support for a tax cut.[199]

With respect to Cohen's broader critique of the use of frequency data to determine individual probabilities when the causal processes involved are unknown or uncertain, the question remains controversial and unresolved, although Cohen's is probably a minority view among probability theorists. The controversy over some difficult hypothetical situations, however, masks the broader points on which there is considerable agreement. First, frequency data is certainly relevant to questions regarding frequentist probabilities, and even Cohen would agree that it is also relevant when it is reasonable to believe that the frequency data is being generated by a relevant causal process.[200] By the same token, when it is reasonable to believe that the frequency data has no relevance to the causal processes involved, consideration of the frequency data would be

197. D. C. Makinson, *The Paradox of the Preface*, 25 ANALYSIS 205, 205–07 (1965); John L. Pollock, *The Paradox of the Preface*, 53 PHIL. SCI. 246, 246–47 (1986).

198. *See infra* Parts III–V.

199. Gregory D. Stanford, *The Average Reader Could Use a Lesson in Math*, MILWAUKEE J. SENTINEL, Feb. 2, 2003, at 4J.

200. That is presumably why Cohen believes the frequency data concerning the diagnostic test is relevant, even if the frequency of the disease in the general population is not—and why he believes that frequency data about incidence of the disease among people who share "relevant characteristics, such as age, medical history, [and] blood group" would be relevant, but not data about those who share first names. Cohen, *Human Irrationality*, *supra* note 172, at 329.

a mistake, even under strict Bayesian principles.[201] This leaves as hard cases the intermediate ones, where the causation involved is unclear or unknown, and where both frequency and nonfrequency evidence are potentially relevant. Here, where theory provides no satisfactory answer, it is prudent to consider each policy question that raises issues of probability judgment on its own terms, in its specific context, and with particularized consideration of the quality of the evidence available and of our understanding of the causal processes involved. That is the project undertaken in the next three sections of this essay.[202]

D. Behavioral Theory and Legal Policy

Behavioral theory first entered scholarly legal debates as an empirical rebuttal to rational choice models, which assumed, usually on little or no empirical evidence, that individuals could accurately assess risk and other probabilistic concepts, or that errors in such probability assessments would be randomly distributed and could therefore be ignored.[203] Policy prescriptions were often based on such rational choice assumptions. Behavioral theory poses a serious challenge to such work because it demonstrates that individuals often do not assess frequentist probabilities in accordance with the assumptions of rational choice theory, and that these errors are not random, but may be skewed in a particular direction (such as that of availability) and shared by a majority of the population.[204]

Behavioral theory therefore does pose a powerful counterargument to the assumptions of rational choice and to policy arguments that rely on rational choice models. At the very least, behavioral theory would caution against using rational choice models to formulate policy without a substantial empirical inquiry into how people actually assess risk and probability in specific legal and policy contexts. Legal theorists, however, have in recent years attempted to go beyond mere critique to build positive theoretical arguments and policy prescriptions that rely, in various ways, on the insights and findings of behavioral theory.[205] After all, if behavioral economics shows that people deviate from rationality in consistent and predictable ways, then should not policymakers take those failures of rationality into account in formulating legal rules, and do their best to counter any harm caused by such irrational biases and heuristics?

201. Cohen's bilharzia example, *supra* note 196, would be an example of such a causally irrelevant frequency, although one might argue, with Blackburn, that it is not relevant "because we can envisage a different, narrower reference class" under which the individual probability judgment should be put (for example, the frequency of bilharzia among people who don't wade in fresh water). Blackburn, *supra* note 194, at 331.

202. *See infra* Parts III–V.

203. *See* RICHARD A. POSNER, ECONOMIC ANALYSIS OF LAW §§ 1.1–2.3 (3d ed. 1986).

204. *See, e.g., supra* Part II.A.

205. *See, e.g.,* Symposium, *Empirical Legal Realism: A New Social Scientific Assessment of Law and Human Behavior*, 97 NW. U. L. REV. 1075 (2003).

The previous two sections of this essay have highlighted some of the dangers of this approach. First, while behavioral theory may show that individuals do not make decisions on a purely rational basis, the heuristics and biases that are the subject of study by behavioral economists do not show consistent "errors," "underestimations" or "overestimations," of frequency or probability. Indeed, the underlying evolutionary assumption of behavioral economics is that these heuristics developed (and were retained by the species) precisely because they were useful most of the time in making "correct" judgments of frequency and probability. The problem is not that heuristics such as availability are consistently wrong, but that they are not consistently right. In some circumstances they can, as Kahneman and Tversky showed, lead to widespread and predictable errors. In others, they may lead to quick and relatively accurate assessments of frequency and probability. In yet others, their accuracy may depend on precisely how the question is framed.[206] It is therefore only in specific contexts, particularly those where we have some other more accurate measure of probability, that we can confidently assess whether the availability heuristic leads to overestimation, underestimation, or correct results.

Accordingly, any analysis of the effects of heuristics such as availability must first consider the nature of the probability judgment involved. In cases where the probability judgments call for frequencies, and those frequencies can be compared with reliable statistical data on the same or similar frequencies, we can determine with some confidence whether the use of heuristics leads to incorrect results.[207] In cases where the probability judgments call for subjective probability assessments of individual events, we can be far less confident in determining the accuracy of such judgments. Rather, the issue for policymakers in such cases will often be very similar to the debate about the application of Bayes's Theorem described in the prior section: whether it is appropriate to use frequentist data, even if weak or relating to poorly understood causal processes, to critique individual subjective probability judgments. In cases where we think it appropriate to criticize individual probability judgments, an important subsidiary question is whether the incorrect judgments can be altered by additional or alternative presentation methods.

The next three sections analyze some ongoing policy debates that have used behavioral theory in attempting to describe and assess peoples' probability judgments concerning the issues involved. In accordance with the philosophical discussion of probability in the previous sections, these debates have been divided into three categories: (1) those

206. See the discussion of Kahneman's and Tversky's experiment on "representation vs. availability" *supra* notes 141–52 and accompanying text.

207. *See, e.g.*, Paul Slovic et al., *Rating the Risks*, *in* THE PERCEPTION OF RISK 104, 111–19 (Paul Slovic ed., 2000) (demonstrating variance between expected and actual frequency of causes of death).

issues primarily involving the effect of policy on groups of people, such as environmental and toxicological hazards, where use of frequentist data to correct mistakes in probability judgments is common and appropriate;[208] (2) those issues involving individual assessments of risk, such as products liability, where frequentist data has been used to critique individuals' assessments of their own risks, but where the use and policy implications of such data are more controversial;[209] and (3) those issues involving individual probability assessments, such as contract performance or litigation results, where there is little or no basis for preferring frequentist data to subjective individual probability assessments.[210]

III. HAZARD REGULATION AND RISK PERCEPTION: THE CASE OF SUNSTEIN V. SLOVIC

It is a well-documented and significant fact that many of the things Americans fear most are not the things that cause them the most harm.[211] They overestimate the fatalities associated with motor vehicle accidents, pregnancy, tornadoes, floods, cancer, and homicides, while underestimating the fatalities relating to smallpox vaccinations, diabetes, lightning, stroke, tuberculosis, and asthma.[212] These are errors in the estimation of frequencies of precisely the sort studied by Kahneman and Tversky, and seem explicable, at least in significant part, by the availability heuristic.[213] Death by car accident, childbirth, and homicide are all vivid and easily recalled. Scenes of devastation caused by floods and tornadoes are frequently seen on television. Diabetes, stroke, tuberculosis, and asthma, in contrast, are generally associated with debilitation rather than death, and are less likely to be topics of either news reports or conversations among friends and acquaintances.

These frequency errors have policy implications. If people are mistaken about the fatalities associated with various activities, then they are likely to favor overexpenditure of funds to prevent damage from hazards such as floods and tornadoes, while underfunding efforts to reduce diseases such as diabetes and asthma, which they view as less dangerous. Accordingly, these cognitive errors can lead to bad and distorted policymaking. Perhaps the strongest exponent of this critique of risk regulation has been Cass Sunstein.[214] In a series of recent articles, Sunstein,

208. *See infra* Part III.
209. *See infra* Part IV.
210. *See infra* Part V.
211. *See, e.g.*, Slovic et al., *supra* note 207, at 111–19.
212. *Id*. at 107.
213. *Id*. at 105.
214. CASS R. SUNSTEIN, AFTER THE RIGHTS REVOLUTION: RECONCEIVING THE REGULATORY STATE 92 (1990); Kuran & Sunstein, *supra* note 153, at 735–46; Richard H. Pildes & Cass R. Sunstein, *Reinventing the Regulatory State*, 62 U. CHI. L. REV. 1, 60–61 (1995); Cass R. Sunstein, *Behavioral Analysis of Law*, 64 U. CHI. L. REV. 1175, 1187–90 (1997); Cass R. Sunstein, *Health-Health Tradeoffs*,

along with a number of coauthors, has focused on the cognitive aspects of risk regulation, and has tried to develop a theory of the regulatory state that recognizes that peoples' perceptions of risk will often be distorted by heuristics and biases.[215] He is also sensitive to the way risks are presented in the media and in public fora, and warns of the effect of "informational cascades" (repeated vivid retellings of memorable stories and occurrences), which enhance the effect of the availability heuristic.[216]

Sunstein's work seems to offer the clearest and most coherent use of cognitive theory to demonstrate that many of the demands on policymakers (as well as their responses) are not only irrational, but based on mistaken probability judgments. This is true in large part because Sunstein focuses not on individuals' perceptions of their own personal risks, but of the riskiness of various hazards as a basis for policymaking.[217] Policymaking, by its nature, is an activity of general import, and when people—either experts or lay people—speak of potential risks in connection with government policy, they are concerned with more than their own personal safety. An individual may be so afraid of sharks that he never swims in the ocean, thereby reducing his personal risk of shark attack to zero, yet may still overestimate the danger of shark attacks, probably because he sees many reports of shark attacks on summer television news programs. It is his generalized perception of the frequency of shark attacks that is wrong. This is why Sunstein can be so confident, from a frequentist probability analysis, that most people overestimate the danger of shark attacks.[218]

Indeed, when I first planned this essay, I intended to use Sunstein's work as a relatively unproblematic example of the use of behavioral theory in policy analysis, because it seemed to me unassailable that cognitive biases and errors can be shown in the estimation of the frequency of injury from various hazards. I still believe that to be the case as far as it goes, but the debate over risk perception and its effect on policymaking has taken an interesting and somewhat metaphysical turn. It has become, at least in part, a debate over the objective reality of risks.

Sunstein's leading opponent in this debate is Paul Slovic, a psychologist who has studied risk perception for many years and was a coeditor, with Kahneman and Tversky, of their famous book on judgment under uncertainty.[219] Whereas Sunstein, the lawyer, emphasizes the tendency of people to make objectively verifiable cognitive errors, Slovic, the cognitive scientist, emphasizes the inherent subjectivity of risk perception and the rationality of many intuitive approaches to risk. Slovic

63 U. CHI. L. REV. 1533, 1538–60 (1996); Cass R. Sunstein, *Which Risks First?,* 1997 U. CHI. LEGAL F. 101, 115–21; Sunstein, *supra* note 53, at 163–68.

215. *See* sources cited *supra* note 214.

216. Kuran & Sunstein, *supra* note 153, at 685.

217. *See* sources cited *supra* note 214.

218. *See* Sunstein, *supra* note 53, at 1134.

219. HEURISTICS & BIASES, *supra* note 72.

found, for example, that although people made errors in estimating the fatalities associated with various hazards, their perception of the riskiness of such activities varied far more from that of experts than their mistakes concerning fatalities.[220] In fact, lay persons' estimates of risks of various hazards differed significantly from their own estimates of the fatalities attributable to such hazards.[221] In short, people viewed riskiness as something other than the frequency of deaths caused by such activities. A prime example of this was nuclear power.[222] When asked to consider the risk of dying from thirty various activities, most lay people ranked nuclear power as the highest risk.[223] Experts ranked it number twenty.[224] Yet when asked to estimate the actual fatalities attributed to these risks, nuclear power had both the lowest fatality estimate and highest perceived risk.[225] The primary reason for this, as Slovic notes, is the disaster potential of nuclear power—the belief most people have that a nuclear accident would result in enormous numbers of casualties, far more, in fact, than the experts believe would occur.[226] Most of the other hazards on this list, while not having the same disaster potential as nuclear power, still elicited risk perceptions that varied according to more than just the number of fatalities they were believed to generate.[227] Other qualitative features that Slovic found made people view activities as riskier were whether the activities were involuntary, delayed, uncontrollable, dreaded, or severe (certainly fatal).[228] Also, products and activities that were viewed as having substantial benefits, as well as harms, were seen as less risky than those with few benefits.[229]

These studies led Slovic to conclude that there is a strong subjective element to risk perception.[230] In part, this is because such perception involves considerations of values as well as facts. Because most people disapprove of nuclear weapons, warfare, DDT, handguns, crime, and nuclear power, they see them as greater risks than swimming pools, home appliances, and downhill skiing, even if the latter cause more deaths to Americans in the average year.[231] Such determinations are surely not "wrong" in any factual sense, but represent the individualized value judgments of citizens in a democratic society. But Slovic makes a much

220. Slovic et al., *supra* note 207, at 113–16.
221. *Id*.
222. *Id*. at 115.
223. *Id.*
224. *Id.*
225. *Id.*
226. *Id.* at 117. More than forty percent of respondents expected more than 10,000 fatalities in a "disastrous year" for nuclear power. Twenty-five percent expected more than 100,000 fatalities. Government experts predicted that the "maximum credible nuclear accident" would involve only 3300 "prompt fatalities." *Id*.
227. *Id.* at 114–18.
228. *Id.* at 117.
229. *See id*. at 118–19.
230. *Id*. at 119.
231. *Id*. at 115.

stronger statement. He sees subjectivity in the very concept of risk itself.[232] As he states:

> The probabilities and consequences of adverse events are assumed to be produced by physical and natural processes in ways that can be objectively quantified by risk assessment. Much social science analysis rejects this notion, arguing instead that risk is inherently subjective. In this view, risk does not exist "out there", independent of our minds and cultures, waiting to be measured. Instead, human beings have invented the concept of risk to help them understand and cope with the dangers and uncertainties of life. Although these dangers are real, there is no such thing as "real risk" or "objective risk." The nuclear engineer's probabilistic risk estimate for a nuclear accident, or the toxicologist's quantitative estimate of a chemical's carcinogenic risk, are both based on theoretical models whose structure is subjective and assumption-laden, and whose inputs are dependent upon judgment.[233]

This is too much for Sunstein.[234] Although he finds much to praise in Slovic's book, denying the reality of risk altogether seems to him nonsensical.[235] While agreeing that risks can be described and studied in different ways, and that science may not always provide clear answers to questions of risk, he forcefully defends the reality of the concept of risk.[236] He states:

> I am sure that Slovic would agree that smoking three packs of cigarettes per day produces a real risk, one that is both objective and high, whether or not it can be quantified with precision. Slovic would also agree that swimming at the beach in Marblehead, Massachusetts, produces a real, but small, risk of being killed by a shark even if we are unable to quantify it. Of course, we can describe these risks, consistent with the evidence, in many different ways. But this point does not establish that risks are not real or that they are not objective. Indeed, Slovic agrees that "the dangers and uncertainties of life . . . are real." If dangers are real, so are risks.[237]

Sunstein and Slovic both seem to believe their disagreement is primarily about the degree of deference to give to expert risk assessment. But in fact their argument is over a deeper philosophical issue—one that is central to this essay. They are arguing about the reality of probability. Slovic's statement that "there is no such thing as 'real risk' or 'objective risk'"[238] is reminiscent of a famous statement by Bruno De Finetti, one of the founders of the subjectivist concept of probability, who declared that

232. Paul Slovic, *Trust, Emotion, Sex, Politics and Science, in* THE PERCEPTION OF RISK, *supra* note 207, at 390, 392–93.

233. *Id*. (citations omitted).

234. Sunstein, *supra* note 53, at 1147 (citations omitted).

235. *Id*.

236. *Id*.

237. *Id*.

238. *See supra* text accompanying note 233.

"probability does not exist."[239] His followers are quick to point out that what he meant was that "objective probability does not exist."[240] While expressed as a metaphysical statement, the conclusion is more easily defended on epistemic grounds. There is no objective concept of risk to individuals because there is no objective method for verifying or refuting statements of individual risk.

Consider Sunstein's statement that "swimming at the beach in Marblehead, Massachusetts, produces a real, but small, risk of being killed by a shark."[241] Now imagine an individual who actually swims at Marblehead. One of two things will happen: Either he will be killed by a shark, or he will not. Which of those two events will prove that he had "a small but real risk" of being killed? Being killed may prove it was a risk, but surely not that it was small one. Not being killed proves neither that a small risk existed nor that it did not. The concept of a "small risk" only appears when we consider the risks to a group or in other frequentist terms. If tens of thousands of people swam at Marblehead over the last five years and only one was killed by a shark, then we might say that the risk of shark attack at Marblehead generally, or to the average or random swimmer, was small.

By the same token, consider Slovic's statement that both "the nuclear engineer's probabilistic risk estimate for a nuclear accident" and the "toxicologist's quantitative estimate of a chemical's carcinogenic risk" are based on theoretical models, the structure of which is "subjective and assumption-laden."[242] This statement ignores significant differences in the objective evidence likely to support those two different risk assessments. The nuclear engineer's risk assessment is likely to be nothing more than a subjective probability judgment of the likelihood of nuclear accident by someone with specialized knowledge of nuclear power reactors. There is insufficient evidence of actual nuclear accidents to compute a reliable base rate, and the likelihood of a repeat of Chernobyl or Three Mile Island is itself a matter of subjective probability assessment. Moreover, such assessments, even when made by experts, are subject to numerous heuristics and biases. The toxicologist's quantitative estimate of a chemical's carcinogenic risk, in contrast, is likely to be based primarily on frequency distributions of cancer in groups exposed to the chemical—individuals, experimental animals, or both. To be sure, this data may be inconclusive and subject to interpretation, yet a toxicologist's estimate of carcinogenic risk, well-supported by quantitative empirical data, will be entitled to greater "weight" as a probability judgment than even the most careful risk assessment by the nuclear engineer. This is because, as Cohen would argue, we believe these studies to be directly

239. BRUNO DE FINETTI, THEORY OF PROBABILITY x (1979).
240. DECISION, PROBABILITY AND UTILITY, *supra* note 62, at 97.
241. Sunstein, *supra* note 53, at 1147.
242. Slovic, *supra* note 232, at 392–93.

relevant to the processes that will cause cancer in humans, whereas no nuclear expert can know the cause of the next nuclear accident.

Although Sunstein and Slovic seem to be arguing over whether to trust experts or lay persons' assessments of risk, the real question is *when* to trust experts and when the weight to be accorded expert's probability judgments is insufficient to justify such trust, particularly when it is at variance with lay peoples' perceptions of risk. In general terms, when the risk involved appears subject to accurate assessment through frequentist methodologies, deferral to expertise is more appropriate than when the experts merely offer a subjective risk assessment, which may differ from that of lay people. Some of the factors that may make frequentist methodologies more appropriate are the existence of a series or group of similar, repeatable individual events or occurrences, the ability to accurately assess results with respect to the entire group or a representative sample of it, and a belief that natural processes are acting on all the members of the group in a similar way and are likely to act in the same way in the future. In short, science is not equally knowledgeable about all kinds of risks, and not all kinds of risks are equally amenable to scientific inquiry. Sunstein is right that availability cascades and cognitive error may often lead to lay risk assessments that are contrary to sound scientific data, but Slovic is also right that subjectivity, uncertainty, and cognitive biases may play a dominant role in expert opinions that are not entitled to be privileged over those of lay persons. The question is one that must be decided case-by-case, by scientists and policymakers debating amongst themselves. When scientists tell us that in recent years there have been no abnormal rates of injury or death attributable to eating apples sprayed with Alar, we can reasonably conclude that the risks of Alar have been exaggerated. When experts point out that in recent years there have been no abnormal rates of injury or death due to nuclear reactors, we can equally reasonably conclude that this provides no basis for changing our personal assessment of the risk of such reactors.

IV. PRODUCTS LIABILITY RISK AND REGULATION: DO YOU FEEL LUCKY?

Behavioral studies have also had a profound impact on recent debates over products liability law.[243] Many of these debates have centered around the issue of "enterprise liability"—the claim that, at least in part due to the cognitive limitations of product users, companies should be

243. Steven P. Croley & Jon D. Hanson, *Rescuing the Revolution: The Revived Case for Enterprise Liability*, 91 MICH. L. REV. 683, 706–12 (1993); Richard A. Epstein, *The Unintended Revolution in Product Liability Law*, 10 CARDOZO L. REV. 2193, 2202–05 (1989); George L. Priest, *The Invention of Enterprise Liability: A Critical History of the Intellectual Foundations of Modern Tort Law*, 14 J. LEGAL STUD. 461, 517 (1985); Alan Schwartz, *The Case Against Strict Liability*, 60 FORDHAM L. REV. 819, 832 (1992).

held liable for all injuries caused by the products they manufacture.[244] An extensive theoretical literature has developed as to whether such rules are normatively justified, with both proponents and opponents agreeing that a critical question is whether consumers can and do accurately assess the risks of the products they purchase.[245] If consumers do make accurate assessments of product risk, the argument goes, then market forces will provide the correct amount of safety in manufactured products. If, however, consumers systematically underestimate the risks of the products they buy, then enterprise liability will be justified to ensure the optimum level of safety.[246] The question of enterprise liability, then, hinges at least in part on a cognitive question: Do consumers accurately perceive the risk of the products they buy, or do they systematically over- or underestimate them?

Not surprisingly, behavioral studies have been enlisted on both sides of this argument. Some scholars, citing studies of availability, representativeness, optimism, and avoidance of cognitive dissonance, argue that people are likely to underestimate the risk of most ordinary products they buy.[247] Others, citing availability in particular (accidents are vivid and easy to recall), as well as risk aversion, argue that it is more likely that people overestimate most product risks than underestimate them.[248] Overestimation of risk is a problem that is exacerbated rather than solved by enterprise liability.

Many observers of this debate have recognized that behavioral theory does not currently provide either sufficiently clear and determinate empirical evidence or a sufficiently powerful theory of human behavior to resolve these disputes.[249] But there is a deeper theoretical problem with this debate, related to the nature of probability judgments themselves. When legal theorists argue over whether consumers over- or underestimate product risks, what is the correct measure of risk to compare against consumer opinion? The obvious answer, "the true risk of such products," is fraught with difficulties.

244. While many commentators recognize the doctrine of strict products liability as a move in the direction of enterprise liability, most current proponents of enterprise liability envision a much more comprehensive regime that would make companies essentially insurers for all injuries caused by the products they sell. *See, e.g.*, James A. Henderson, Jr. & Jeffrey J. Rachlinski, *Product-Related Risk and Cognitive Biases: The Shortcomings of Enterprise Liability,* 6 ROGER WILLIAMS U. L. REV. 213, 225–55 (2000).

245. *See generally* Croley & Hanson, *supra* note 243, at 716–17 (discussing the effects of consumer risk analysis on notions of product liability).

246. *See id*. at 707–08; *see also* Epstein, *supra* note 243, at 2205 (noting that enterprise liability "makes sense" if one believes that consumers lack ability to prevent losses); Priest, *supra* note 243, at 527 ("The unavoidable implication of the three presuppositions of manufacturer power, manufacturer insurance, and internalization is absolute liability."); Alan Schwartz, *Proposals for Product Liability Reform: A Theoretical Synthesis*, 97 YALE L.J. 353, 374 (1988) (noting that "strict liability may be justified . . . if the [risk information] assumption—that consumers know risks of harm—is false").

247. Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: Some Problem of Market Manipulation,* 74 N.Y.U. L. REV. 630, 696–704 (1999) [hereinafter Hanson & Kysar, *Problem*].

248. *Id*. at 704–14.

249. *Id*. at 721–23.

The "true" risk of such products surely cannot mean the objective risk determined on a frequentist basis. Product risk can vary greatly from person to person, based primarily on the skill of the user and the uses to which it is put. The risk of accidents from motor vehicles, for example, is substantially greater for teenagers than for older Americans.[250] Product purchases are individual purchases, and the market model, by assuming that people are able to ascertain and purchase products with the risk characteristics they desire, requires that the true assessment of risk be the risk posed by that product to the individual consumer. This, of course, raises the now-familiar problem of subjective probability assessment. If a person believes that his risk is lower than average because he is a better than average driver (as most people do),[251] then there is no empirical evidence that can prove such a person wrong.

But this is a bit too easy and, in the context of products liability, even a touch disingenuous. Although we cannot directly refute individuals' subjective assessments of the risk particular products pose to them, we can raise doubts about their assessments indirectly in a number of ways. First, we can compare many peoples' subjective risk assessments with the objectively determined frequency risk to determine whether there is an aggregate over- or underestimation of individual product risk. If most people believe that their individual risk from using the product is less than the average frequentist risk of harm, then it is plausible to claim that most people are underestimating the risk of the product. Plausible, but surely not provable. Some of the people assessing themselves to be at low risk may be considering risk factors that actually do reduce their likelihood of harm. Others, however, may simply be reflecting optimism or other heuristic biases. Some may be doing a little of both. A simple comparison of aggregate risk assessments gives us no way to determine the accuracy of any individual risk assessment. Indeed, there is no such way. With hindsight, most products will injure a relatively small number of people (whose "true" likelihood of injury may be said to be one, or certainty). Risk only enters the picture as a counterfactual claim about what might have happened. Yet aggregate underestimation of individual risk relative to objective frequentist risk, if it could be shown, might be enough to justify policy arguments in favor of enterprise liability.

A stronger justification, however, would be to show that cognitive errors occur, not only at the aggregate level, but at the individual level; that individuals make mistakes about their own individual risks of harm. This cannot be done with aggregate statistics alone, but requires an analysis and critique of the individual risk assessment process itself, a cri-

250. *See* NAT'L HIGHWAY TRAFFIC SAFETY ADMIN., TRAFFIC SAFETY FACTS 2002: YOUNG DRIVERS 1–2 (2002), *available at* http://www-nrd.nhtsa.dot.gov/pdf/nrd-30/NCSA/TSF2002/2002ydrfacts.pdf (last visited June 7, 2004).

251. *See* Ola Svenson, *Are We All Less Risky and More Skillful Than Our Fellow Drivers?*, 47 ACTA PSYCHOLOGICA 143 (1981).

tique similar to the one Kahneman and Tversky made in their psychiatrist hypothetical.[252] If it can be shown that individuals systematically overvalue some risk factors (like those most vivid and available), while slighting or ignoring others (like relevant prior probabilities), then departure from rational choice, and from Bayesian models of decision making, can be shown at the individual level. This is more easily done with some product risks than others. It is hard to say whether a frequent driver who has never been in an accident, and rates her risk from driving as substantially below average, is truly skillful or merely lucky. That is because we do believe, based on empirical studies, that different kinds of drivers, with different skill levels and other characteristics, have different risks of accidents.[253] With respect to other types of risks, however, we believe, again based on the relevant science, that individual characteristics play a less important role. Accordingly, we would be much more likely to say that a frequent smoker who rated her risk of cancer as substantially below average was making a cognitive error than we would the experienced but frequent driver who claimed a below-average risk of accident. Smoking risk has therefore become a somewhat paradigmatic case for demonstrating cognitive errors in subjective risk assessment.[254]

There is a substantial literature on individuals' risk assessments of the dangers of smoking.[255] Broadly speaking, it shows that:

(1) Most individuals, including both smokers and nonsmokers, substantially overestimate the frequency of lung cancer caused by smoking. In a famous study, W. Kip Viscusi found that when people were asked: "Among 100 cigarette smokers, how many do you think will get lung cancer because they smoke?" The mean response was forty-three.[256] Viscusi estimated the true risk, based on frequency of actual illness among smokers, to be between five and ten.[257] This is also consistent with more general studies of risk assessment, which show that relatively low probability risks with available outcomes tend to be overestimated.[258]

(2) When smokers are asked to state their own subjective probability of contracting lung cancer from smoking, their assessments of their personal risk is substantially below that of the frequentist estimates in

252. *See supra* notes 159–65 and accompanying text.

253. There is much empirical evidence to indicate, for example, that beginning drivers are at greater risk of accidents than average, as are very old drivers and those who regularly drive in excess of the speed limit. *See supra* text accompanying note 250. Given the high correlation of accidents with drunk driving, it is also likely that teetotalers have a lower risk of accidents than those who frequently go out and drink at parties, restaurants, and bars. *See* Sharon E. Conaway, *The Continuing Search for Solutions to the Drinking Driver Tragedy and the Problem of Social Host Liability*, 82 NW. U. L. REV. 403, 403 (1988).

254. *See infra* text accompanying notes 256–67.

255. *See infra* notes 256–67 and accompanying text.

256. Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: Some Evidence of Market Manipulation,* 112 HARV. L. REV. 1420, 1503–04 (1999) [hereinafter Hanson & Kysar, *Evidence*].

257. *Id*. at 1504.

258. *Id*. at 1503.

the Viscusi study.[259] For example, Hanson and Kysar discuss and cite various studies that indicate "that smokers perceive smoking as significantly less risky for themselves than for other smokers, that smokers view their own risks as not significantly higher than those for [nonsmokers], and that smokers tend to underestimate the actual risks to themselves."[260] Slovic similarly cites numerous studies indicating that "strong optimism biases have been found in cigarette smokers," and concludes that "[y]oung smokers are highly likely to judge themselves as less at risk from cigarettes than they judge the 100 hypothetical smokers asked about in Viscusi's survey questionnaire."[261]

(3) Perhaps the most striking results of these studies, however, is their demonstration of the malleability of subjective risk assessments and the importance of context in asking for and making such assessments. For example, Slovic and others have been able to show that asking for a quantitative statement of the probability of a single outcome (like Viscusi's question about lung cancer) results in higher probability judgments than questions that ask for a comparison of probable outcomes.[262] When Slovic asked how many of one hundred smokers were likely to die from each of fifteen different potential causes of death, the projected incidence of lung cancer deaths went from a mean of fifty-six (in a Viscusi-style single outcome question), to a mean of twenty.[263] There is also substantial evidence that most people are not very good at using absolute mathematical probability statements to indicate their subjective beliefs about risk. This may be, in part, the result of using an unfamiliar language to talk about risk. American basketball players, for example, might give wildly inaccurate estimations of the height of a basketball hoop in meters, but that would be a reflection of their unfamiliarity with the metric system, not unfamiliarity with a basketball court.[264] Similarly, studies that find substantial disparities between an individual's quantitative probability assessment and probabilities expressed verbally may reflect most people's difficulties in using numbers to express probabilistic concepts.[265] Other studies, cited by Hanson and Kysar, showed that ordinal rankings of various risks (most to least dangerous) were more consistent and less subject to framing effects than absolute estimates of numerical risk.[266] People also frequently make the mistake noted in

259. *Id*. at 1511–16.

260. *Id*. at 1511–27.

261. Paul Slovic, *Do Adolescent Smokers Know the Risks?*, *in* THE PERCEPTION OF RISK, *supra* note 207, at 366.

262. *See id*. at 366–67.

263. *Id.* at 367.

264. Paul D. Windschitl, *Judging the Accuracy of a Likelihood Judgment: The Case of Smoking Risk*, 15 J. BEHAV. DECISIONMAKING 19, 21 (2002).

265. *See* Paul D. Windschitl & Gary L. Wells, *Measuring Psychological Uncertainty: Verbal Versus Numeric Methods*, 2 J. EXPERIMENTAL PSYCHOL.: APPLIED 343, 357–58 (1996).

266. Hanson & Kysar, *Evidence*, *supra* note 256, at 1519. The ability of most people to judge relative risk more accurately than they can compute mathematical probabilities lends some support to

William Safire's predictions at the beginning of this essay, assigning a range of mutually exclusive possible events an aggregate probability of greater than 100.[267]

The malleability and diversity of these effects in producing wildly different probability estimates for smoking risk has led some behavioral theorists to a more radical conclusion than simply positing that various heuristics cause people to make cognitive errors.[268] Some theorists are now suggesting that people may not really be making any subjective probability judgments at all.[269]

Remember that the theory of subjective probability began as an attempt (primarily by philosophical economists) to give empirical content to probability judgments applied to single nonrepeat events.[270] Although such statements had no empirically verifiable truth conditions, they were said to correspond to psychological facts about an individual's beliefs, beliefs that could be ascertained, at least in theory, by an individual's betting behavior.[271] An individual who believed an event was eighty percent likely to occur should be willing to make a 1-4 bet in favor of its occurrence, or a 4-1 bet against. Bets (and therefore probability judgments) that do not conform to this calculus enable a "Dutch book" to be made against you, and are considered irrational, even in the realm of subjective probability.[272] Behavioral research has shown, however, that people make such irrational probability judgments quite frequently.[273] This may be due to an unfamiliarity with math, or with mathematical probability, but it may also reflect a deeper rejection of the very premises on which subjective probability theory is based.

Paul D. Windschitl has recently stated that "[a]n underrecognized property of subjective probability judgments is that they are typically *ad hoc* constructions."[274] In support of this view, he cites various studies of framing and context effects.[275] Perhaps most striking is a study by Slovic and Monahan in which respondents were given vignettes of mental health cases and asked to judge the probability that the individuals in the vignettes would do harm to others.[276] When asked to judge that probability at ten point intervals on a scale from zero to one hundred, the mean

Cohen's claim that people judge probability in Baconian, not Pascalian terms. *See supra* notes 86–89 and accompanying text.

267. *See, e.g.*, Amos Tversky & Derek J. Koehler, *Support Theory: A Nonextensional Representation of Subjective Probability*, 101 PSYCHOL. REV. 547, 553–54 (1994).

268. *See infra* notes 274–83 and accompanying text.

269. *See infra* notes 274–83 and accompanying text.

270. *See supra* Part I.B.

271. *See supra* Part I.B.

272. *See supra* note 62.

273. *See supra* note 62.

274. Windschitl, *supra* note 264, at 25.

275. *Id*.

276. Paul Slovic & John Monahan, *Probability, Danger and Coercion: A Study of Risk Perception and Decisionmaking in Mental Health Law*, *in* THE PERCEPTION OF RISK, *supra* note 207, at 347.

probability of doing harm was .44.[277] When the response scale was changed to add six response categories less than ten, the mean probability given by respondents dropped to .12.[278] When the study was repeated with mental health professionals, similar results were obtained.[279]

The notion that subjective probability judgments are ad hoc and easily malleable are also supported by Slovic's findings about the affect heuristic.[280] Slovic notes that "[w]ithin an analytic view of judgment and decision-making, risk and benefit are distinct concepts."[281] A substantial body of research, however, has shown an inverse relation between risk and benefit.[282] In general, people tend to view activities they like and approve of as being less risky than those they view negatively.[283] In short, it appears likely that how much we like or approve of something may also affect our probability judgments concerning it.

The ad hoc nature of subjective probability judgments should not really surprise us. Such statements are, after all, introspective reflections of our perception of our own uncertainty, with no empirical verifiability and little opportunity for learning or improvement. Accordingly, we are unlikely to be certain about our own degree of uncertainty, rendering such judgments easily susceptible to change with the occurrence or report of vivid events (availability), with different formulations of potential questions and answers (framing effects), and even with changes in our likes and dislikes (the affect heuristic). But the ad hoc nature of subjective probability judgments does pose a serious problem for policymakers. Under rational choice models, individuals were believed to make predictable, consistent, and accurate (on average) probability judgments, and to act in accordance with them.[284] Under behavioral economic models, individuals were believed to make predictable, consistent, and *inaccurate* probability judgments, and to act in accordance with them.[285] Under this new ad hoc theory of subjective probability judgments, individual probability judgments are not predictable, frequently inconsistent, and do not govern subsequent action.[286] Rather, probability judgments are strongly influenced by and vary substantially according to the particular circumstances under which they are made, and are developed as part of a decision-making process, rather than prior to it.

277. *Id*. at 352.

278. *Id*.

279. On the response scale with ten point intervals, twenty percent of mental health professionals rated the probability of causing harm at .10 or less. When the scale with additional responses less than ten was used, the percent of respondents making a probability judgment of less than .10 rose to 49.3%. *Id.* at 356.

280. *See* Melissa L. Finucane et al., *The Affect Heuristic in Judgments of Risks and Benefits*, *in* THE PERCEPTION OF RISK, *supra* note 207, at 415.

281. *Id*.

282. *Id*.

283. *Id*.

284. *See supra* text accompanying note 203.

285. *See supra* Part II.

286. *See supra* notes 268–83 and accompanying text.

This new aspect of behavioral theory has just begun to affect policy arguments in those areas where sophistication about behavioral theory is greatest.[287] For example, in the products liability field, recent scholarship seems to be shifting away from a debate between those who believe heuristics lead to systematic overestimation of risks and those who believe it leads to systematic underestimation.[288] Rather, it is becoming a debate over how the law should deal with the malleability and ad hoc nature of the perception of product risk by most consumers.[289] Broadly speaking, this is a debate between the educators and the "protectors."[290]

The educators are those who believe that the best way to deal with most people's difficulties in evaluating product risk is by providing clear, vivid, comprehensible, and accurate information about those risks.[291] While recognizing that no campaign of product risk information will be received, understood, and acted upon by everyone, they point to the responsiveness of most people to various warnings—particularly to salient and specific warnings—as evidence that people's risk assessments can be made more accurate, and that this should be a primary goal of policymakers.[292] One leading advocate of this approach is W. Kip Viscusi.[293] While acknowledging all the evidence (including his own) that people frequently over or under assess relevant risks, Viscusi insists that, "[n]otwithstanding the difficulties individuals often have in processing risk information, hazard warnings can be a potentially effective tool in altering individual perceptions and influencing risk-taking decisions."[294] After discussing various warning labels and information-provision campaigns that have been successful in reducing risky behavior (including smoking), and explaining which features of warnings make them particularly effective, Viscusi recognizes that no amount of warnings and information will enable everyone to make an accurate assessment of risk.[295] He argues that the decision whether to rely on warnings or more "protectionist" measures should be governed by a cost-benefit analysis.[296] As he writes:

> In general, while there will be some percentage of failures in the population, we cannot target our interventions to those specific people. Instead, we should ask whether the warnings pass muster for the entire market and, if so, then warnings should remain the desired policy approach even though direct control of the product

287. *See infra* notes 291–301 and accompanying text.
288. *See infra* notes 291–301 and accompanying text.
289. *See infra* notes 291–301 and accompanying text.
290. *See infra* notes 291–301 and accompanying text.
291. *See infra* notes 293–97 and accompanying text.
292. *See infra* notes 293–97 and accompanying text.
293. W. Kip Viscusi, *Individual Rationality, Hazard Warnings, and the Foundations of Tort Law*, 48 RUTGERS L. REV. 625, 636 (1996).
294. *Id*. at 650.
295. *Id*. at 650–67.
296. *Id*. at 667–69.

> risk characteristics might protect the small segment of consumers who do not make sound decisions.[297]

Critical to Viscusi's policy approach, therefore, remains his assumption that most people can be educated to make "sound decisions," which are presumably those based on accurate probability assessments. The various studies he cites of warnings that have been fairly effective in changing consumer behavior make his conclusion possible, but by no means necessary.

On the other side of this debate are Jon Hanson and Douglas Kysar, who, in a series of articles, have reached entirely different policy conclusions also based, in part, on the malleability and ad hoc nature of probability judgments.[298] Hanson's and Kysar's central point is that risk assessments are not only manipulable, but that there are powerful and sophisticated market actors with the ability and incentive to manipulate them—namely product manufacturers.[299] They argue that under the current products liability regime, manufacturers do manipulate consumer perceptions, generally in the direction that causes consumers to ignore or underestimate the risks of the manufacturers' products.[300] The result, they argue, is that a significant number of consumers are manipulated into irrational underassessment of risk, thereby justifying a regime of enterprise liability.[301]

It might seem that the evidence of the malleability of probability judgments has simply spawned a new version of the rationality/irrationality debate, with Viscusi and others arguing that most consumers can be warned and informed into a state where they understand and appreciate the relevant risks, and Hanson and Kysar contending that a significant number will inevitably, under current market conditions, be manipulated into underestimating the relevant product risks. Such a characterization of the debate would be encouraging, because it makes it

297. *Id*. at 668.

298. Hanson & Kysar, *Problem*, *supra* note 247; Jon D. Hanson & Douglas A. Kysar, *Taking Behavioralism Seriously: A Response to Market Manipulation*, 6 ROGER WILLIAMS U. L. REV. 259 (2000) [hereinafter Hanson & Kysar, *Response*]; Hanson & Kysar, *Evidence*, *supra* note 256.

299. Hanson & Kysar, *Evidence, supra* note 256, at 1572; Hanson & Kysar, *Problem, supra* note 247, at 633–40.

300. Hanson & Kysar, *Response*, *supra* note 208, at 387–88. Hanson and Kysar do not try to quantify the extent of this underassessment, but they clearly view it as significant enough to warrant massive expenditures by product manufacturers. As they write:

> [M]anufacturers and marketers in consumer product markets can shape consumer risk assessments by altering the way they manufacture, package, and market their products. What is more, that possibility creates profit-enhancing opportunities for manufacturers and marketers by affording them a way of increasing consumer willingness to pay. As a result, the problem of market manipulation seems inescapable in an unregulated consumer product market. Manufacturers, to survive, *must* behave "as if" they are attempting to manipulate consumer risk perceptions. And in light of the immense power of the market forces driving these attempts, it seems highly doubtful that manufacturer strategies (be they deliberate or accidental) will fail.

Hanson & Kysar, *Problem*, *supra* note 247, at 747.

301. *See supra* note 300.

appear that the disagreement is largely a factual one, and subject, in principle at least, to empirical determination.[302]

But there are more radical implications to the malleability of probability judgments, insofar as this research casts doubt on the basic premise that consumer decisions are purely or even primarily the result of cognitive processes at all, whether accurate or inaccurate. Professors Henderson and Rachlinski, at the end of an article analyzing the Hanson-Kysar position,[303] come close to raising these questions:

> If consumer preferences are completely constructed, then what exactly is supposed to be the efficient level of consumption? Should the socially optimal demand for soup be measured with the cans in alphabetical order, or not? On a rainy day, or sunny? With what kind of music or ambient odors (if any) in the background? In what section of the store? What should the labels look like? How big are the cans? Risk is no different. The slight risk of death from skiing creates part of the sport's pleasure whereas the slight risk of death from exposure to a nearby hazardous waste dump creates a massive uproar. The notion that manufacturers distort consumer risk-perception assumes that there is some natural and appropriate risk-benefit assessment from which manufacturers lead consumers astray. If we take seriously the psychological proposition that all preferences are constructed, then there is no magical correct level of risk that consumers should endure.[304]

Henderson and Rachlinski correctly note the socially constructed nature of both preference and risk, but they also conflate some important differences between the two. Economics generally assumes that consumer preferences are exogenous and does not seek to apply concepts of rationality to demand for consumer goods, be they tulips, hula hoops, or Pokemon cards.[305] They do, however, assume rationality in seeking to satisfy those preferences.[306] That is, someone who values a Pokemon card at more than $100 should be willing to pay $100 for such a card. Conversely, anyone willing to pay $100 for such a card may be said to value it at least that much. Risk is different. When someone purchases a product, its riskiness is not an objective fact common to all users, but is unique to that user of that product. Nonetheless, we can sometimes calculate an approximation of that risk using Bayesian concepts, particularly when we believe the risk is relatively invariable among different individuals, to arrive at something we might call the correct or accurate risk perception. Indeed, it is the very possibility of speaking of an accurate

302. Hanson and Kysar seem to support this view when they write, "As may now be obvious, one's position on the question of how well the existing law addresses the problem of manipulation depends upon how significant one perceives the underlying problem to be." Hanson & Kysar, *Response, supra* note 208, at 297.

303. Henderson & Rachlinski, *supra* note 244.

304. *Id*. at 258.

305. Hanson & Kysar, *Problem, supra* note 247, at 636.

306. *See supra* note 203 and accompanying text.

risk perception that makes it possible for us to talk about over- or underestimation of risk. Nonetheless, the only way we can perceive individual risk assessment is by looking at individual consumption of risky products. It is here that the paradox emerges.

Consider an individual who pays $1000 extra for side airbags in his car, has a radon detector in his house, and smokes a pack of cigarettes a day. Is there any way we can determine whether that individual is underestimating the risk of smoking or simply has a strong preference for smoking despite the risk? Recent work on behavioral theory indicates that this may well be an unanswerable question.[307] Risk perceptions do not precede purchasing decisions, but are made in conjunction with them. We can, of course, ask smokers what they think their risk of contracting illness is, and we are likely to get relatively high estimates that vary greatly depending on precisely the way the question is asked. We can also observe behavior and see many people acting inconsistently with regard to risks, taking great precautions with respect to some risks while indulging in others. Such behavior is compatible with cognitive errors and lack of accurate information about risks, but it is equally compatible with different preferences and values.

The fundamental problem is that products liability, in particular, is an area of probability where neither objective nor subjective concepts predominate. The choices involved are sufficiently individual and personalized that we recognize individual preferences will determine the purchase and use of risky products. But unlike other features of products, such as color or size, where we assume that individual preference is the only relevant factor, there are enough objectively determinable facts about product risk that someone who acts or speaks in a way that ignores those risks appears to be making a cognitive error, not merely expressing an unusual preference. Recent behavioral research indicates that we will not be able to resolve this feeling/cognition dichotomy with respect to probabilities, and that perceptions of risk and feelings toward risk are inextricably linked in human beings.[308] Accordingly, any normative theory that depends on viewing probability judgments either entirely as a matter of preferences or entirely as a matter of cognition will be unsatisfactory.

What is required is a normative theory recognizing and incorporating the linkage of cognition and affect in the field of product risk. Implicit in both Viscusi's and Hanson-Kysar's work are the normative bases for such theories. For Viscusi, this entails an emphasis on the greater choice, product innovation, and individual responsibility that would result from a fault-based liability regime emphasizing warnings and availability of information, whether or not such warnings were effective in

307. *See* Paul Slovic, *Introduction and Overview, in* THE PERCEPTION OF RISK, *supra* note 207, at xxxii.

308. *Id*.

promoting "rational" product choice and use by most consumers.[309] The availability of the information and potential freedom of individuals to make rational choices could be sufficient to justify such a regime. After all, neither democracy nor market capitalism are justified by the claim that people make correct choices, but by the freedom to choose. In this view, it is not necessary to resolve the unanswerable question whether the purchase of a risky product is based on risk preference or cognitive error. It is a normative view that privileges choice over protection.

For Hanson and Kysar, it involves emphasizing the incentive effects on manufacturers in a regime of enterprise liability.[310] Because manufacturers would be responsible for all costs of accidents, they would seek to reduce the objective frequency of risk to the greatest extent possible consistent with the maximization of sales and earnings.[311] While this might mean that some risky products could no longer be sold legally (for example, hang gliders and cigarettes), because the objective frequency of injury from such products is likely to exceed the earnings generated by their sales, Hanson and Kysar would presumably view this not as a problem, but as a cost-effective solution to objective product risk. Here again, it does not matter whether consumer demand for the product was based on risk preference or cognitive error. This is a normative view that privileges protection over choice.

While this may seem to be replacing an intractable factual dispute with an intractable normative dispute, I would suggest that the issue can perhaps be dealt with more easily on a product-by-product basis. It may well be the case that a societal consensus will emerge in favor of product choice and innovation in certain areas (such as recreational equipment) and in favor of protection and strict enterprise liability in others (such as drugs). The reasons for such emerging attitudes may be complex, involving a whole series of interrelated variables about the way various products are used and perceived.[312] Yet if, as I have argued, the dispute over

309. Viscusi notes that "[w]arnings take advantage of the heterogeneity of individual responses to risk and enable people to make choices consistent with their own preferences," but adheres to the preference/cognition dichotomy under which the "benefits" of "avoiding mistaken decisions" must be weighed against the "cost" of "elimination of choices in the marketplace." Viscusi, *supra* note 293, at 668.

310. Hanson and Kysar already argue that enterprise liability creates an

> enhanced incentive for manufacturers to monitor and manage the rate and severity of accidents involving their products. A product risk that goes undiscovered by consumers and for which manufacturers are not liable under a fault-based regime may well turn out under enterprise liability to be one that manufacturers identify and prevent or, alternatively, disclose and explain to consumers, all in an effort to lower liability costs.

Hanson & Kysar, *Response*, *supra* note 208, at 304–05.

By viewing risk prevention and explanation of risks as "alternatives" presumably to be chosen by the manufacturer on the basis of whichever more effectively reduces objective risk and therefore liability, Hanson and Kysar imply that manufacturers should not concern themselves with subjective risk preferences.

311. *See supra* note 310.

312. See, for example, the complex interrelationship of factors described in THE PERCEPTION OF RISK, *supra* note 207, at 152.

products liability regulation is essentially a normative one, albeit one frequently involving objective degrees of risk, any potential resolution of the issue must rest on existing or emerging social norms.

The Strange Case of Divorce

Another legally significant area where neither frequentist nor subjective concepts of risk predominate is divorce law. Virtually everyone knows the famous statistic that fifty percent of all American marriages end in divorce.[313] It seems to be the case, however, that virtually everyone who gets married, despite their knowledge of the statistic, believes that theirs will be one of the fifty percent of marriages that will succeed.[314] From a purely statistical standpoint, this is a major cognitive error, with people ascribing high probabilities of success to their marriage prospects when the true frequentist rate of success is much lower. This certainly looks like overconfidence or the optimism heuristic at work.

Yet we tend not to believe that divorce is the result of some relatively uniform or stochastic process, but rather of complex and individual decisions of which true causes are known, if at all, only to those with detailed knowledge of the particular failed marriage. We have no less an authority than Tolstoy to tell us that each unhappy family is unhappy in a unique way.[315] Aren't divorces, and predictions concerning the probability of divorces, the kind of unique and individual events where subjective assessments by the parties are a better source of individual probabilities of marriage success than frequentist probability? In most cases of individual and uniquely caused events, I would agree, but it seems to me that predictions about divorce by those about to be married represents a strange and subtly exceptional case.

First is the fact that, from a frequentist perspective, the error rate is huge. Almost half the people who predict the success of their marriages are going to be wrong.[316] Despite the fact that the risk of divorce is common and well-understood, and that people are aware of its frequency in American marriages generally, people underestimate it enormously when applied to their own potential marriages. This leads us to a second concern—the quality of the underlying subjective probability assessments. It seems plausible that people in the process of applying for marriage licenses are not in the best position to make a dispassionate analysis of the factors that might lead to their divorces at some time in the future.

313. ROSE M. KREIDER & JASON M. FIELDS, U.S. CENSUS BUREAU, NUMBER, TIMING, AND DURATION OF MARRIAGES AND DIVORCES: 1996 (Feb. 2002), *available at* http://www.census.gov/prod/2002pubs/p70-80.pdf (last visited June 7, 2004).

314. Lynn A. Baker & Robert E. Emery, *When Every Relationship Is Above Average: Perceptions and Expectations of Divorce at the Time of Marriage*, 17 LAW & HUM. BEHAV. 439, 443 (1993).

315. LEO TOLSTOY, ANNA KARENINA 3 (Nina Berberova & Leonard J. Kent eds., Random House 1965) (1878).

316. *See* Baker & Emery, *supra* note 314, at 443.

Indeed, they seem very likely to commit the kind of cognitive errors Kahneman and Tversky warned against.[317] Rather than ask, "How frequently do people with characteristics like ours (young and in love) ultimately wind up divorced?"—a question likely to lead them to the base rate of fifty percent—they are more likely to try to compare the characteristics of their relationships with those of people who are actually divorced, or to try to construct "scenarios" under which they would no longer want to stay married, and find the characteristics far different or the scenarios implausible. Finally, recognizing the close relationship between emotion and cognition, these predictions of perpetual marital bliss may just as well be viewed as expressing something other than a cognition—a hope, desire, or an expression of determination to beat the odds.[318]

Unsurprisingly, commentators on divorce law policy have viewed these data as demonstrating that many people fail to accurately predict the probability of their own divorce.[319] They differ, however, just as the products law scholars do, over whether people can be informed or induced into making more rational risk assessments.[320] Yet the differences appear not to be as stark as in the products debates. Elizabeth Scott, who generally favors making divorce more difficult for couples that sign a premarital contract to that effect, argues that such precommitments are likely to induce people considering entering a marriage relationship to think about it more rationally.[321] She also argues that such people are likely to have a long-term preference for stability, which may be thwarted by cognitive errors or short-term preferences leading to divorce.[322] Yet she recognizes, to the extent behavioral theory implies that most people cannot be induced to "calculate more carefully" their marriage decisions, that there are "limitations in the applicability of the model."[323]

Professors Lynn Baker and Robert Emery, in contrast, view their results as demonstrating that "legal reforms aimed at increasing rational ex ante responsiveness to divorce statutes are not likely to succeed."[324] Unlike Hanson and Kysar, however, they see this not as a result of manipulation by any third party, but simply of the "systematic optimism" of people contemplating marriage.[325] Like Hanson and Kysar, however, they argue that well-crafted default rules, not ex ante rational choice, are

317. *See supra* Part II.

318. In this regard it is perhaps relevant that the respondents in the Baker and Emery study were people who had recently applied for a marriage license. Baker & Emery, *supra* note 314, at 440.

319. *See infra* notes 321–26 and accompanying text.

320. *See infra* notes 321–26 and accompanying text.

321. Elizabeth S. Scott, *Rational Decisionmaking About Marriage and Divorce*, 76 VA. L. REV. 9, 57–59 (1990).

322. *Id*. at 14–21.

323. *Id*. at 69.

324. Baker & Emery, *supra* note 314, at 448.

325. *Id*.

the only way to ameliorate the deleterious consequences of mistakes in risk assessment.[326]

My purpose in this section has not been to resolve these complex policy disputes, or even to inflict my views on matters about which others are far more knowledgeable. Rather, it has been to try to demonstrate why these problems are so difficult. Unlike debates about objective frequency risks, where Sunstein and his colleagues can simply demonstrate that anyone who thinks that shark attacks are a bigger problem than diabetes is wrong, it is not the case that anyone who thinks they are at less risk than average from smoking is necessarily wrong. If too many people believe it, some of them will undoubtedly be wrong. Whether people can be taught, induced, or persuaded not to make such mistakes, or whether mistakes are the inevitable result of giving people choices, is the deep and perhaps unanswerable question that pervades these policy debates.

V. CONTRACTS AND CLAIMS: HOW TO ACT WHEN THE RISKS ARE BASELESS

A third category of legal issues involves risks and probabilities where frequencies and base rates are either nonexistent or of very limited utility in calculating individual probabilities. In these situations, a competent individual's own subjective assessment of the risk involved is likely to be the best probability judgment available. Accordingly, determining the accuracy of such probability judgments, except for testing them against certain minimum rationality constraints, is simply not possible. Yet even here, some policy debates have been thought to turn on the unanswerable question of whether most individuals are accurately assessing the relevant risks.[327] Law and economics theorists presume such accuracy in accordance with expected utility theory.[328] Opponents cite behavioral studies to demonstrate systematic inaccuracy.[329] Without reliable base rates or risk factors to analyze, however, such debates are at best unresolvable and at worst meaningless.

A. Likelihood of Contract Performance

One area where some such debates have occurred is contract law. Contracts, after all, are in a sense predictions about the future. The risk of nonperformance exists with respect to every contract, and various contract terms are designed to minimize or compensate for that risk. Much of contemporary contract theory assumes that, in negotiating such contracts, the parties accurately analyze and value the risk of nonperform-

326. *Id*. at 448–49.
327. *See* Sunstein, *supra* note 53, at 1152.
328. *See* Larry T. Garvin, *Adequate Assurance of Performance*, 69 U. COLO. L. REV. 71, 140–42 (1998); sources cited *infra* note 348.
329. *See infra* notes 332–37, 353–59 and accompanying text.

ance.[330] In recent years, behavioral theory has been cited in an attempt to question that assumption.[331]

Consider Melvin Eisenberg's use of behavioral concepts to justify what he calls a "dynamic"—individual and discretionary—judicial approach to enforcing liquidated damage clauses.[332] He cites two behavioral heuristics that he believes lead contracting parties to systematically underestimate the risk of liquidated damages provisions. One is optimism, which Eisenberg says will cause a contracting party to "believe that his performance is more likely, and his breach less likely, than is actually the case."[333] The other is availability, which Eisenberg says "may lead a contracting party to give undue weight to his present intention to perform, which is vivid and concrete, as compared with the abstract possibility that future circumstances may compel him to breach."[334] He goes on to write:

> Because a contracting party is likely to take the sample of present evidence as unduly representative of the future, he is apt to overestimate the extent to which his present intention to perform is a reliable predictor of his future intentions. Because actors have faulty telescopic faculties, a contracting party is likely to overvalue the benefit of the prospect of performance, which will normally begin to occur in the short term, as against the cost of the prospect of a breach, which will typically occur, if at all, only down the road. Because actors tend to underestimate risks, a contracting party is likely to underestimate the risk that a liquidated damages provision will take effect.[335]

Professor Larry Garvin invokes behavioral concepts similar to Eisenberg's in his article on adequate assurance of performance, but is more equivocal about the implications of these concepts to the accuracy of assessments of the risk of nonperformance.[336] After describing availability, overoptimism, prospect theory, and framing, he concludes, "Cognitive psychology, dispelling the overly optimistic assumptions of expected utility theory, thus shows that the promisee is unlikely to evaluate risk accurately at the time of contracting, and will usually err on the side of underestimating infrequent risk."[337]

One can certainly quarrel with these authors' conclusions that behavioral heuristics will usually lead to underestimation of nonperformance risk. Aren't prior instances of contract breach likely to be vivid and easily recalled (like business failures by friends and relatives)? If so,

330. *See* Garvin, *supra* note 328, at 140–42; sources cited *infra* note 348.
331. *See infra* notes 332–37.
332. Melvin Aron Eisenberg, *The Emergence of Dynamic Contract Law*, 88 CAL. L. REV. 1743, 1784 (2000).
333. *Id*.
334. *Id*.
335. *Id.*
336. Garvin, *supra* note 328, at 140–70.
337. *Id*. at 161.

then shouldn't the availability heuristic lead to their overestimation? Is optimism truly a cognitive error, or an expression of hope in the face of uncertainty? Because breach is a low-probability risk in most instances,[338] aren't Eisenberg and Garvin improperly ignoring the evidence that most low-probability risks are overestimated?

One can raise all these issues, but they seem to pale in light of the most basic question: Compared to what are these risks being "underestimated"? Both Eisenberg and Garvin assume that the frequency of contract breach is very low. Garvin provides some data from credit sales indicating default rates of about two percent.[339] Even if a true rate could be computed, however, the overall frequency of contract breaches would surely not provide an accurate statement of the risk of breach for any particular contract. We have no basis for believing that the individual risk of nonperformance in all or most such cases is remotely similar to that hypothetical average rate. Indeed, we believe precisely the opposite. Unlike disease etiology, where our best current scientific understanding is that most people who engage in similar practices (such as heavy smoking) are subject to similar risks, we make precisely the opposite assumption with respect to contract formation. We assume that the likelihood of nonperformance varies greatly depending on the nature of the contract, the business, goods, or services involved, and the resources, intentions, and relationships of the contracting parties. We believe these factors are far more important than the fact of contracting in determining the risk of nonperformance.[340] Unless these beliefs are shown to be false, there is simply no basis for the claim that contracting parties systematically underestimate their risks of nonperformance.

Such a refutation is theoretically possible. Consider the previous discussion of divorce.[341] Marriage is a contract with a rather high risk of breach when computed on a frequentist basis.[342] Yet the contracting parties, according to recent studies, almost always rate their own risks of breach as exceedingly low.[343] Because approximately fifty percent of

338. As Garvin writes, "After all, the great majority of contracts are performed in full, making breach exceptional." *Id.* at 148.

339. *See id.* at 148 n.360.

340. There are of course institutions with an interest in compiling statistical data on frequency of contract breach and relevant risk factors. Banks and other credit-granting institutions generally maintain such data. Yet even if we assume that an individual is contracting with an institution with good statistical data about the frequency of breach among people of his or her age, marital status, and income level, is that institution likely to be able to make a better assessment of the risk of that individual's likelihood of breach than the individual? Generally, this is the situation in which economists worry about adverse selection, because the individual's greater knowledge about his or her personal situation is presumed to create an information asymmetry enabling the individual to make the more accurate assessment of risk. *See, e.g.,* Kenneth S. Abraham, *Environmental Liability and the Limits of Insurance*, 88 COLUM. L. REV. 942, 946–47 (1988); Christopher A. Richardson, *The Community Reinvestment Act and the Economics of Regulatory Policy*, 29 FORDHAM URB. L.J. 1607, 1616–17 (2002).

341. *See supra* text accompanying notes 313–26.

342. *See supra* notes 313–14 and accompanying text.

343. *See supra* notes 313–14 and accompanying text.

these people will in fact get divorced,[344] it is not unreasonable to say that these people were mistaken in subjectively viewing their risks of divorce as low, and that frequentist measures of divorce risk are more accurate than the subjective views of the contracting parties.

In contractual situations where the frequency of breach is much lower, however, it is far more difficult to demonstrate the inaccuracy of subjective risk assessment. Assume that the statistical rate of breach with respect to a certain category of contract can be shown to be five percent, but that contracting parties consistently rate their own risks of breach at only one percent. Even leaving aside the problem of using mathematical concepts to express low-probability events, do these facts indicate general inaccuracy in subjective probability assessment? After all, ninety-five out of one hundred people who believed they had a very low risk of breach will feel themselves to have been proved right. The other five are likely, in hindsight, to view their risks of nonperformance as having been much greater than five percent. An objective observer, reviewing in hindsight the individual facts that caused the nonperformance, is also more likely in most cases to agree with the high subjective probability instead of the low frequentist probability.[345]

Even more than in the products field, policymaking with respect to contractual performance cannot depend on unresolvable questions about the accuracy or inaccuracy of individual risk assessments. All we can truly say about the real probability of contractual breach is that most people who enter into contracts intending to perform will do so. A relatively small number will not, and will view themselves, with hindsight, as having made a mistake about their probabilities of performance. Both Professor Eisenberg's argument for a dynamic approach to liquidated damages and Professor Garvin's proposal for adequate assurance of performance clauses can, and indeed must, rest on the observation that a small percentage of contracting parties will fail to perform in accordance with their intentions. Ascribing the cause of such breaches to a general "cognitive bias" adds no more to the argument than ascribing them all to "fate" or "bad luck." A hypothetical statistically derived "true" rate of contract breach has no more reality than those other concepts.

B. Probability of Success in Litigation

Finally, we come to the paradigmatic legal case of probability judgment: assessing the chances of success in an individual litigated case. Unlike the situation with environmental risks, products, or contracts, no commentator has claimed that lawyers are systematically making substantial errors in their estimates of the probability of success in individual

344. *See supra* notes 313–14 and accompanying text.

345. This may of course be an example of "hindsight bias," or it may be a better assessment based on more complete evidence. In this situation, it is impossible to know which.

cases. Rather, almost all the commentary assumes that lawyers value claims, in substantial part, based on the likelihood of success of that claim, and do so fairly accurately in the great majority of cases.[346]

The reason for this sanguine attitude is not that lawyers are assumed to have greater predictive powers than other experts. Rather, it is based on the indisputable empirical observation that most cases settle.[347] The standard economic model of settlement states that cases will settle whenever both parties to the litigation substantially agree as to the expected value of the claim.[348] This creates a settlement window whereby both parties can save litigation costs by settling for the expected value of the claim. Agreeing on the expected value of the claim, in turn, requires substantial agreement by the parties (or their lawyers) as to the likelihood of success of the claim. Because most cases settle, lawyers must be in substantial agreement as to the probability of success of most cases. There may be many reasons for such agreement, but the commentators generally assume that the settlement agreement is the result of accurate assessment of the litigation risk by both parties.[349] Cognitive biases are referred to only to explain the comparatively rare instances in which the settlement model breaks down, either because settlement does not occur or because the parties settle "frivolous" claims, with negative expected value, for positive amounts of money.[350]

This work has expanded the traditional rational choice model of settlements. Commentators have cited information asymmetries,[351] prospect theory (both risk aversion and risk preference),[352] and asymmetric

346. *See* sources cited *infra* note 348.

347. *See* POSNER, *supra* note 203, § 21.5.

348. *See id.* § 21.5 (analyzing the disposition of civil cases based on the probability of success and the costs of litigation and settlement); William F. Baxter, *The Political Economy of Antitrust*, *in* THE POLITICAL ECONOMY OF ANTITRUST: PRINCIPAL PAPER BY WILLIAM BAXTER 13 (Robert D. Tollison ed., 1980); Robert H. Mnookin & Lewis Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce,* 88 YALE L.J. 950, 973–77 (1979); I. P. L. P'ng, *Strategic Behavior in Suit, Settlement, and Trial*, 14 BELL J. ECON. 539, 544 (1983); George L. Priest & Benjamin Klein, *The Selection of Disputes for Litigation*, 13 J. LEGAL STUD. 1, 13–17 (1984); Steven Shavell, *Suit, Settlement, and Trial: A Theoretical Analysis Under Alternative Methods for the Allocation of Legal Costs*, 11 J. LEGAL STUD. 55, 56–57 (1982); Donald Wittman, *Dispute Resolution, Bargaining, and the Selection of Cases for Trial: A Study of the Generation of Biased and Unbiased Data*, 17 J. LEGAL STUD. 313, 320–22 (1988); Charles M. Yablon, *The Good, the Bad and the Frivolous Case: An Essay on Probability and Rule 11*, 44 UCLA L. REV. 65, 75 (1996).

349. *See* sources cited *supra* note 348.

350. *See infra* notes 353–59 and accompanying text.

351. *See* Lucian A. Bebchuk, *Suing Solely to Extract a Settlement Offer*, 17 J. LEGAL STUD. 437, 437–41 (1988) (writing that defendants may not be able to judge the strength of plaintiffs' claims, and that it may be expensive for defendants to develop this information); Avery Katz, *The Effect of Frivolous Lawsuits on the Settlement of Litigation*, 10 INT'L REV. L. & ECON. 3, 4–5 (1990).

352. Janet Cooper Alexander, *Do the Merits Matter? A Study of Settlements in Securities Class Actions*, 43 STAN. L. REV. 497, 530–31 (1991) (writing that defendants consider some cases too big to lose); John C. Coffee, Jr., *Understanding the Plaintiff's Attorney: The Implications of Economic Theory for Private Enforcement of Law Through Class and Derivative Actions*, 86 COLUM. L. REV. 669, 702 (1986); Chris Guthrie, *Framing Frivolous Litigation: A Psychological Theory*, 67 U. CHI. L. REV. 163, 168 (2000) (using prospect theory to predict that "plaintiffs pursuing low-probability claims are likely

stakes by the parties in the litigation to explain these phenomena. But, unlike other policy areas involving risk assessments, commentators have generally not posited the existence of systematic mistakes by lawyers in judging the likelihood of success of litigation claims.

Chris Guthrie comes closest to making such an claim when he tries to explain how frivolous cases can be brought and settled for amounts greater than their true expected values.[353] He proposes that plaintiffs might be "risk seeking" with respect to low-probability claims, and that one of the reasons for this might be "cognitive error" on the part of plaintiffs, induced perhaps by a "salience or vividness heuristic" causing plaintiffs "to recall reported instances of large trial verdicts in apparently frivolous suits, and therefore systematically overestimate their chances of success."[354] By positing an asymmetric risk aversion on the part of defendants, Guthrie can explain why these cases settle, but the fact of settlement implies that the degree of plaintiff overoptimism in most of these cases cannot be too great.[355] One may also question whether plaintiffs' attorneys suffer from the same availability heuristic bias, because they presumably do not get most of their information about the legal system from newspaper reports of large trial verdicts. This raises again the question of framing effects and of whether an attorney who has a lower estimate of the claim's likelihood of success than the plaintiff she is representing can bring about a settlement.[356] Guthrie recognizes, however, that the evidence for cognitive bias is anecdotal and offers it as one of four possible explanations for risk seeking by plaintiffs, some or all of which combine with others.[357]

to prefer the risk-seeking option—trial—while defendants are more likely to prefer the risk-averse option—settlement").

353. Guthrie, *supra* note 352, at 202–04.

354. *Id*. at 202–03.

355. For example, imagine a plaintiff with a claim for $100,000. In reality, the claim has a ten-percent chance of success, for an expected value (EV) of $10,000. If we posit plaintiff litigation costs of $10,000, then this case is technically frivolous, because the expected recovery, after litigation, is zero. The plaintiff, however, due to cognitive bias, mistakenly believes the claim has a thirty-percent likelihood of success, so he mistakenly brings the claim and will accept no settlement under $20,000 (EV - litigation costs). The defendant, on the other hand, correctly values the claim at $10,000, but also faces $10,000 in litigation costs, and because of risk aversion, asymmetric stakes, or some other reason, will pay another $5000 to make the case go away. So defendant offers $25,000 and the case settles. Notice, however, that if plaintiff's optimism became much greater, say fifty-percent chance of success, the case would not settle. Because settlement is the norm, any risk seeking or optimism bias by plaintiffs must be relatively small.

356. In a prior article, Russell Korobkin and Chris Guthrie described experimental evidence that suggested lawyers and clients "systematically evaluate litigation options differently" and that "lawyers are more likely to explicitly or implicitly employ expected financial value calculations," whereas clients are more subject to framing effects and other cognitive heuristics. Russell Korobkin & Chris Guthrie, *Psychology, Economics, and Settlement*, 76 TEX. L. REV. 77 120–25 (1997). They concluded that the "extremely high actual rate of settlement" is likely due in large part to lawyers' ability to persuade clients to settle in cases where they would otherwise have been unwilling to do so. *Id.*

357. Guthrie, *supra* note 352, at 206.

Others have cited the optimism bias as an explanation for why certain cases fail to settle.[358] Again, however, any such optimism bias must be weak, asymmetric, and nonsystematic, because it is clearly not preventing settlement in the vast majority of litigated cases.[359]

Yet the prevalence of settlement surely does not prove that lawyers are generally accurate in their assessments of the probability of success in pending cases. At most, it suggests that lawyers do not disagree dramatically about their probability assessments in most cases. This may be because they are all accurately perceiving the true probability of success of most cases, but it may also be because these lawyers simply share many of the same beliefs and assumptions about what factors make a case "stronger" or "weaker"—even if those beliefs and assumptions are not objectively correct.[360] For example, most lawyers assume that plaintiffs in medical malpractice and products liability claims have a greater likelihood of success in jury trials than in bench trials, even though the statistical evidence of win rates in litigated cases shows precisely the opposite result.[361]

The assumption that lawyers accurately assess the likelihood of success of individual claims is surely open to question, but it is hard to see how that question can be resolved. The reason is that there is no good frequentist data against which to judge those subjective probability judgments. Because the vast majority of cases settle, we can only compute frequencies of success with regard to the relatively small proportion of cases that remain and are ultimately adjudicated. This might be sufficient if the adjudicated cases were a representative sample of all cases brought to court, but there are good reasons to doubt that.

Theodore Eisenberg, who has analyzed and thought about litigation statistics as much as anyone, points out that the cases that proceed to adjudication are unusual in that the parties have been unable to reach a settlement prior to trial or other adjudication.[362] It is unlikely that this fail-

358. OREN BAR-GILL, THE SUCCESS AND SURVIVAL OF CAUTIOUS OPTIMISM: LEGAL RULES AND ENDOGENOUS PERCEPTIONS IN PRE-TRIAL SETTLEMENT NEGOTIATIONS (Harvard Law Sch. Pub. Law, Working Paper No. 35, May 2002), *available at* http://papers.ssrn.com/abstract=318979 (last visited June 7, 2004); Richard Birke & Craig R. Fox, *Psychological Principles in Negotiating Civil Settlements*, 4 HARV. NEGOT. L. REV. 1, 15–16 (1999).

359. Indeed, Bar-Gill seeks to show that the only optimism that will potentially benefit lawyers is what he calls "cautious optimism," which leads to capturing a greater portion of the settlement window, the amounts saved by settling rather than litigating, but does not lead to breakdown of settlement negotiations altogether. BAR-GILL, *supra* note 358, at 1–3.

360. Much of the legal system proceeds from precisely that assumption. Law students are tested on their ability to distinguish good legal arguments (those with a high probability of success) from bad ones. Frivolous cases (those below some threshold probability of success) may not be brought on pain of sanctions. Plaintiffs' lawyers litigating cases on a contingency fee basis rely not only on their ability to distinguish good cases (high probability of success) from bad, but on their ability to cause other lawyers (defense and insurance counsel) to make similar probability assessments.

361. Clermont & Eisenberg, *supra* note 28, at 1133.

362. Theodore Eisenberg, *Testing the Selection Effect: A New Theoretical Framework with Empirical Tests*, 19 J. LEGAL STUD. 337, 337 (1990).

ure to settle occurs at random.[363] The impact these factors have on the ultimate characteristics of the cases that proceed to trial is known as the "selection effect."[364]

It is possible to compile statistical data on success rates for particular types of litigated cases, and these data are widely available, at least with respect to federal cases.[365] The data show consistent plaintiff win rates substantially below fifty percent for almost all categories of tort cases, and plaintiff win rates well above fifty percent for most contract claims.[366] This could be a reflection of the selection effect, with defendants and their insurers willing to settle all but the weakest tort claims, while contract claims involving legitimate business disputes usually settle, with only the most egregious breaches going to trial. On the other hand, this disparity might reflect a difference in the relative strength of all such cases filed. Personal injury plaintiffs, under the contingent fee system, may be willing to file even relatively weak cases, while contract plaintiffs, engaged in ongoing business dealings with defendants, may be unwilling to sue unless the other side's conduct is particularly outrageous.[367] Win rates alone cannot resolve these questions, which require detailed analysis of settled as well as litigated cases.

Furthermore, even if win rates of litigated cases were a good indicator of the probability of success for the *average* case, that would still not make them a better predictor of success in individual cases than lawyers' subjective probability assessments. The reason, once again, is our understanding of the causal factors that generate litigation outcomes. Although there are some stochastic elements involved in the litigation process (jurors and judges are randomly assigned to cases), and some uniform aspects (in general, the same rules of law are applied in similar cases), we tend to believe that the most important determinants of litigation outcomes are individual aspects of the case, which are best known to the parties and lawyers involved. Such individual elements would include the facts, disputed and undisputed, the evidence the parties are

363. For a theoretical discussion of the factors that are likely to prevent settlement of litigation, see Bruce L. Hay & Kathryn E. Spier, *Settlement of Litigation, in* THE NEW PALGRAVE DICTIONARY OF ECONOMICS AND THE LAW 442–45 (Peter Newman ed., 1998); sources cited *supra* note 348. For an empirical investigation of such factors, see Leandra Lederman, *Which Cases Go to Trial?: An Empirical Study of Predictors of Failure to Settle*, 49 CASE W. RES. L. REV. 315, 327 (1999).

364. Eisenberg, *supra* note 362, at 337.

365. For example, win rates of federal civil cases are maintained, based on data compiled by the Administrative Office of the United States Courts, on a user-friendly website by Professors Theodore Eisenberg and Kevin Clermont. For fiscal year 2000, the most recent year available, the plaintiff win rate for all contract cases was .8986. The plaintiff win rate for all personal injury tort cases was .2939. Theodore Eisenberg & Kevin M. Clermont, *Judicial Statistical Inquiry Form*, *at* http://teddy.law.cornell.edu:8090/questata.htm (last visited June 7, 2004).

366. *Id*.

367. Professors Clermont and Eisenberg note that even when the selection effect is present, there may still be a "residual" correlation between win rates and case strength. *See* Kevin M. Clermont & Theodore Eisenberg, *Do Case Outcomes Really Reveal Anything About the Legal System?*, 83 CORNELL L. REV. 581, 591 (1998).

able to offer in support of their versions of the facts, the specific legal contentions involved, and the legal authorities supporting those contentions. Moreover, even many of the stochastic elements of the process, such as the identities of the judge and jurors, will become determinate facts during the course of the litigation process. Given our general belief that individual cases are decided on the facts and the law, based on arguments and evidence known to the lawyers involved in litigating that case, it is certainly reasonable to believe that the lawyers' subjective probability assessments of the likelihood of success in an individual case are more reliable than statistical probabilities based on broad categories of cases.[368]

Yet there are some data that casts doubt on lawyers' presumed ability accurately to predict litigation results. Samuel Gross and Kent Syverud compared defendants' settlement offers with the actual outcomes of trials.[369] Their data indicated that defendants' settlement offers were consistently below plaintiffs' mean recoveries at trial.[370] These results might be interpreted as showing consistent underestimation by defendants of plaintiffs' likelihood of success at trial. But because settlement offers are not the same thing as probability assessments, and because the actual world of litigation and settlement is more complicated than our thought experiments, the conclusions that may be drawn from the actual data are far more equivocal.

Among the interesting features of Gross's and Syverud's data are (1) the large portion of cases (25.2%) in which no monetary settlement offer was made; (2) the fact that plaintiff prevailed in approximately one-third of the zero-offer cases, with a mean award of $108,265 in all zero-offer cases; and (3) the fact that defendants' offers were always less than the mean award for plaintiff at trial.[371]

368. For example, there is consistent statistical data that the rate of reversals in federal criminal appeals is quite low. Generally, about ten percent of all federal criminal convictions get reversed on appeal. Criminal convictions are usually appealed, so ten percent represents a fairly good base rate for the likelihood of a reversal of a federal criminal conviction. But if an experienced criminal lawyer gives the odds of reversal of a particular conviction as fifty percent, should we view that as overoptimistic, an example, perhaps, of the failure to consider base rates? I don't think so. Our theory of causation tells us that reversals do not occur randomly. Rather, they are correlated fairly strongly with errors or questionable decisions by district courts. It is the consistency of district courts in producing such questionable decisions that gives rise to the consistent frequency of reversals. Accordingly, a trained observer should be able to pick out the individuating characteristics of those decisions most likely to be reversed and reasonably assign higher probabilities to those cases. This does not mean such experts cannot be subject to cognitive biases, but it does mean we have no basis for privileging frequentist probability assessments over subjective expert judgments. And while base rates may be ignored, they may also be incorporated into expert analyses. It is not unusual to hear lawyers talk about how hard it is to get reversals in many circuit courts, or how judges do not like to reverse criminal convictions. These are, in essence, reminders to lawyers of the relevant base rates. Lawyers may still be way off base in their probability assessments. The best indicia of this would not be statistical data, but the views of other lawyers.

369. Samuel Gross & Kent Syverud, *Getting to No: A Study of Settlement Negotiations and the Selection of Cases for Trial*, 90 MICH. L. REV. 319, 343 (1991).

370. *Id*. at 344.

371. *Id*.

Gross and Syverud explain these data primarily as a story of strategic bargaining.[372] While settlement offers may be based on an unbiased estimate of plaintiff's likelihood of success at trial, they surely do not have to be. If a defendant believes that settlement can be obtained for some amount less than the defendant's actual estimate of the plaintiff's likelihood of success at trial, then defendant will offer that lesser amount.

Gross and Syverud recognize that the zero-offer cases require a different explanation.[373] They suggest that these cases represent situations in which defendants (who they assume are generally insurance companies) are seeking some benefit other than settling that particular case for the lowest possible amount.[374] They speculate that such benefits might include (1) inducing dismissal of some cases by counsel who do not wish to incur trial costs; (2) creating greater risk for the class of tort plaintiffs, thus inducing lower settlements in other cases; (3) discouraging plaintiff's counsel from taking additional weak cases; or (4) bringing to trial cases in which defendants are actually seeking to create useful precedents.[375]

While these may be plausible suggestions for why insurance companies might make zero offers in cases they viewed as particularly weak, the relatively high success rate that plaintiffs had in such cases seems to also indicate substantial inaccuracy in predicting litigation results. The fact that plaintiffs prevailed in about thirty-three percent of the zero-offer cases, albeit often with a relatively small damage award, casts doubt on their lawyers' ability to accurately predict results in a substantial number of cases.[376]

What could account for such consistent underestimation? It could be actual cognitive misperception by defense counsel. Optimism, abetted by the well-known tendency of lawyers to be convinced by their own arguments, may induce defense lawyers to believe plaintiffs' cases are weaker than they really are. There may also be institutional aspects of the insurance company-defense counsel relationship that tend to favor the lowest probability assessment of cases.[377]

372. *Id.* at 322.

373. *Id.* at 342–43.

374. *Id.* at 343.

375. *Id*.

376. By the same token, approximately one-third of the cases in which defendants made substantial offers (more than $100,000) resulted in no judgment for the plaintiff. *Id*.

377. In effect, this divergence in interest between insurance companies and their counsel (in-house or outside) represents an agency cost that may interfere with optimal settlement decisions. For an interesting discussion of how agency costs created by divergences in the interests of directors and shareholders may have led to less than optimal settlement behavior in the Pennzoil-Texaco litigation, see Robert H. Mnookin & Robert B. Wilson, *Rational Bargaining and Market Efficiency: Understanding* Pennzoil v. Texaco, 75 VA. L. REV. 295, 315–23 (1989).

Other explanations for low settlement offers might be a desire to show "toughness," always a good quality in a litigator, or the reluctance of each individual to propose a monetary settlement when others are deriding the case as worthless. Finally, the fact that defense lawyers make more money from trying cases than settling them, and may particularly prefer to try cases they are likely to win, may contribute to an institutional bias toward underestimating the relative strength of a weak case.

Thus, as evidence of systematic inaccuracies by lawyers in predicting the probability of litigation outcomes, the data compiled by Gross and Syverud are intriguing, but far from dispositive.[378] The data, particularly with respect to zero-offer cases, are compatible with some level of persistent underestimation by defense counsel of the likelihood of success of low-probability cases. Yet even here, the zero offers are, considered in the aggregate, the weakest cases in the Gross and Syverud study. In that sense, the study is consistent with other behavioral studies demonstrating that people are far better at judging relative probabilities (the weakest cases in the group) than absolute probabilities (those cases with less than a ten-percent chance of success). Moreover, given the very high rate of settlement activity generally, cases in which no monetary offer of settlement is made cannot be very frequent occurrences.

Behavioral theory, along with the data on litigation and settlement rates, give us some grounds for suspecting that lawyers may make systematic misjudgments in estimating the likelihood of success in certain types of cases. Like Kahneman's and Tversky's hypothetical psychiatrist, a defense lawyer faced with weak claims may ignore base rates and ask the wrong question. The lawyer may compare the facts and law of this case with other cases that have resulted in big losses for her clients, concluding that this case does not pose much of a threat and perhaps making a zero-settlement offer. The right question, of course, is what percentage of claims like this nonetheless result in a successful plaintiffs' verdict. The answer to that question is more likely to lead to a small, but potentially acceptable, settlement offer. Recognizing the possibility that lawyers may make systematic errors relative to base rates in subjective assessments of individual claims, however, does not imply that such judgments can or should be made some other way. Given the enormous problems with the statistical data available, and given our belief that individual case characteristics are the primary determinants of litigation results, the use of behavioral theory in this area should be to inform and improve lawyers' individual case assessments, not to supplant them.

VI. CONCLUSION

Not all questions of probability have right answers. Some questions of probability involve the relative frequency of certain characteristics among a group of people or events. In such cases, accurate probability judgments can often be made on an objective and empirical basis. When judging the probability of individual, unique events, however, there are no objective facts that can conclusively determine the accuracy of such

378. In fairness to Gross and Syverud, it should be emphasized that their study does not purport in any way to be a demonstration of legal certainty or predictability. Rather, I have tried to hijack their data for my own purposes in this article. Gross and Syverud's article is an important study of the settlement process, and their discovery of substantial amounts of strategic bargaining among litigants seems to me to be a significant and unassailable finding.

judgments, and the relevance of statistical evidence is unclear and contestable.

Not all use of behavioral heuristics leads to errors. Sometimes the availability heuristic will cause a subject to overestimate the frequency of some characteristic within a group, such as the frequency of English words begin with "R." When judging the frequency or probability of other characteristics, however, such as the frequency of English words beginning with "S" or "M," the availability heuristic is likely to lead to correct results. Whether or not behavioral heuristics lead to errors in probability assessment, however, can only be determined for probability questions that have relatively correct and ascertainable answers.

Accordingly, it is erroneous to assert generally that various behavioral heuristics lead to consistent under- or overestimation of probabilities. As this essay has shown, judging the correctness of probability judgments depends on the precise nature of the probability question at issue, as well as the quality of the statistical and nonstatistical evidence available and the causal processes believed to be involved. Behavioral theory can make important contributions to legal and policy debates, but only if it is applied with a sensitivity to context and to the philosophical issues inherent in all probability concepts.