ON THE DEGREE OF CONFIDENCE FOR ADVERSE DECISIONS

*FREDERICK SCHAUER and RICHARD ZECKHAUSER**

ABSTRACT

In a criminal trial, the prosecution needs to prove its case beyond a reasonable doubt. When an individual is accused of wrongdoing outside the criminal process, as when teachers and politicians are charged with sexual harassment and employees are charged with financial improprieties, people often assume that a similarly stringent standard of proof applies. Yet this transfer from criminal law model to other settings is mistaken. When the value of freedom from incarceration is absent, and other values are present, probabilities of "guilt" less than "beyond a reasonable doubt," perhaps only a mere possibility, are often socially, statistically, and morally legitimate bases for adverse decisions. Relatedly, although sound reasons for the criminal law's refusal to cumulate multiple low-probability accusations exist, the reasons for such refusal are often inapt in other settings. Taking adverse decisions based on cumulating multiple low-probability charges is often justifiable both morally and mathematically.

I. INTRODUCTION

SEVERAL years ago a Vermont police officer was reinstated to his position after his acquittal in court on charges of misdemeanor sexual harassment. His acquittal followed a trial in which the complaining witness, a police department employee, testified that the defendant had physically and verbally harassed her over the course of several months. Four other witnesses, however, each prepared to offer similar testimony about the defendant's harassment of her, were not permitted to testify. The jury, limited therefore to hearing just one complaining witness, found that the

* Schauer is Frank Stanton Professor of the First Amendment, and Zeckhauser is Frank P. Ramsey Professor of Political Economy, both at the John F. Kennedy School of Government, Harvard University. The authors are grateful to Derek Bok, Edgar Kaplan, Louis Kaplow, Steven Kelman, Mark Kleiman, Daniel Steiner, Roger Stern, Geoffrey Stone, and Patricia Wald for highly useful comments on earlier drafts, to the *Journal*'s referee for comments that went far beyond the obligations of the role, and to extremely helpful audiences at the Harvard University School of Public Health, the Harvard University Department of Economics, and the Law and Economics Workshop of the Harvard Law School. We are pleased to acknowledge the support of the National Science Foundation and the Joan Shorenstein Center on the Press, Politics and Public Policy.

[*Journal of Legal Studies*, vol. XXV (January 1996)]
 0047-2530/96/2501-0002$01.50

state had not proved its case beyond a reasonable doubt and returned a verdict of not guilty. Immediately after the acquittal, the police department, which had suspended the officer when the criminal charges were first brought, revoked the suspension and reinstated the officer.

This outcome is typical of a range of cases in which potentially adverse decisions in a *nonjudicial* social or professional context—here, the decision whether to retain the officer—must be made in the face of uncertainty about whether the facts that would support an adverse decision are true. Embedded in this hardly unusual story are thus a host of issues of social epistemology, the inquiry into the effect of normative issues of social justice on the determination of questions of factual knowledge. One issue is that the police department appears not only to have considered the result of the criminal trial as dispositive to its own decision whether to retain an employee but also to have supposed that the probability of guilt necessary to secure a criminal conviction is the same probability that the department should employ in deciding whether to continue the officer's employment. Equally important, because the criminal trial excluded from evidence four other complaints, the police department appears to have set aside that evidence as well, even though most nonjudicial decision makers would have considered the four complaints material to the department's internal decision. Yet the way in which the employment decision tracked the criminal process seems odd; not only does the excluded evidence seem material to determining whether the particular acts charged had occurred, but in an employment context the issue is frequently whether a pattern of behavior exists, rather than whether the defendant committed one particular act. If so, then the "discrete event" approach of the criminal law, in which guilt or innocence is judged for a single act, may ill serve the assessment goals of numerous nonjudicial domains.

We address these two related questions: First, what standard of "proof" is appropriate in taking a decision adverse to an individual, when the adverse decision is something other than a judicial determination of criminal guilt?[1] Second, under what circumstances, outside of the judicial

[1] In ordinary language, to *prove* something is, perhaps, to establish it to a virtual certainty, as when someone accused of some misdeed belligerently claims "you can't prove anything," or when we say that a mathematical theorem has been proved. But in law, the "burden of proof" refers to the *idea* of a level of confidence necessary for a proposition to be taken as established by the evidence, without referring specifically to any *particular* level of confidence. Our use of "prove" and "proof" throughout follows the legal usage and is virtually synonymous with "probabilistic assessment." Thus, to take an extreme case, we do not see it as a linguistic error when someone says that a fact has been "proved to a probability of .20."

system, is it appropriate to take prior acts into account in assessing the probability that a person has committed the act or acts now under investigation? As to the first, we engage the question of when degrees of confidence less than "beyond a reasonable doubt," perhaps far less, may not bar a decision adverse to an individual when the consequences of that decision are other than criminal conviction. The possibility of taking an adverse nonjudicial decision on evidence the judicial system would consider scanty has been prominent in recent years. The issue received attention during the hearings leading to the confirmation of Justice Clarence Thomas, where the opposing forces debated whether the standard for determining the truth of Professor Anita Hill's allegations of sexual harassment should be that of proof beyond a reasonable doubt, as Justice Thomas and his supporters urged, or something dramatically lower, as argued by opponents of the nomination. Similar questions surround the allegations of sexual harassment made against former Senators Brock Adams and Robert Packwood; again, the denial of the allegations by both senators necessitates determining the correct burden of proof when the consequence of being found guilty is censure or expulsion from a political position. Similarly, how strong must the evidence be in order to suspend a professional license? Should Mike Tyson, convicted of rape, be denied a license to continue as a professional boxer? Should William Kennedy Smith, acquitted of rape, be denied a license to practice medicine? Should Tonya Harding, who pled guilty to complicity in an attack on a skating opponent, be permitted to resume a career as a figure skater? How much confidence in the existence of suspected misconduct is necessary when an employer contemplates terminating an employee? And what if the question is which candidate to hire in the first place? How strong should the evidence of misconduct be before someone reports a suspicion in a letter of recommendation, or before a journalist includes it in a newspaper article? How convinced should we be before we no longer buy from a butcher suspected of giving short weight? Even in the most personal and mundane of decisions, we again must determine the burden of proof. Are a few bad line calls sufficient to exclude a tennis player from our next game? How much evidence of racist beliefs, or of pilfering the silver, do we need before we strike someone from the guest list for a dinner party?

As to our second question, the relevance of prior acts in the nonjudicial context has been even less well analyzed. The judicial system, with some exceptions, addresses charges one at a time and is reluctant to consider prior misconduct as relevant to determining the truth of the charge now being considered, but outside the judicial system such skepticism about the relevance of prior acts seems odd. Often the relevant behavior before some nonjudicial decision maker is not a single act, but a pattern of acts.

Outside of the judicial system, how should we think of a series of charges each of which, taken alone, might not satisfy what would otherwise in that context be the necessary probability of guilt? The methods of probabilistic analysis tell us how to combine individual probabilities to produce a compound assessment, yet few guilt-assessing systems are designed to incorporate the mathematics of compounding.

Section II of this article clarifies the idea of "guilt" and the relationship of guilt in the criminal law to guilt in various other contexts. Section III addresses the necessary confidence level in various fact-finding settings, concluding that the criminal law standard of "proof beyond a reasonable doubt" is generally inappropriate outside of the criminal law and that in some contexts probabilities much lower even than .5 may be sufficient to support an adverse decision. Section IV analyzes how multiple low-probability charges might be compounded to support an adverse decision, and Section V suggests a way to operationalize the theory. Noting that systems for combining the probabilities of low-probability charges are in place in a few domains, we offer suggestions for adapting such systems to other domains in which they might be useful.

II. Isolating the Question

In order to focus on the question of degree of confidence, we assume the materiality of the proposition whose truth we are trying to assess to the decision we are proposing to take. It is one thing to ask whether a proposition, if true, would be material to a decision; it is quite another to ask what it is to take the proposition as true for the purposes of the decision. Consider the failure to attend to this distinction in the 1992 presidential campaign, when allegations arose about the sex life of then-candidate Bill Clinton. Almost all who either criticized or supported the decision by the mainstream press to republish allegations first printed in a supermarket tabloid commingled questions about the *validity* of the evidence with questions about the *materiality* of the charges to Clinton's candidacy.[2] Yet it is one issue whether acts of marital infidelity, even if established to a virtual certainty, are material to assessing the qualifications of a candidate for the presidency. Quite another, however, is determining how confident a national newspaper should be of the truth of material charges before publishing them. In 1992, the dispute about the former inhibited clear thought about the latter, since those who ques-

[2] We use "materiality" in its technical sense. Nontechnically, the word "relevance" combines the question of the degree to which a proposition is established with the question of whether a proposition, if established, is germane to the decision to be made. In the law of evidence, relevance refers only to the former and materiality only to the latter.

tioned the reliability of the evidence often simultaneously articulated doubts about the materiality of the charges.

To avoid this confusion, we limit our analysis to allegations of acts that are plainly material to the decision at hand.[3] Thus, we exclude cases where materiality is juxtaposed with epistemology: for example, charges of marital infidelity in a political context (although charges of infidelity would be plainly relevant in selecting a marriage counselor or spouse), charges that a job applicant committed a crime 30 years ago, or charges that a candidate for the Baseball Hall of Fame gambled on baseball games. In contrast, true charges that a candidate for state treasurer embezzled corporate funds, that a sitting judge takes bribes, that a babysitter is a child molester, that a used car dealer turns back odometers, that a store employee pilfers from the inventory, or that a candidate for tenure plagiarized his publications would all undeniably justify an adverse action that would not be taken were the charges false.

In some of these examples, the conduct is treated by the law as a crime, even though in the contexts that concern us here the decision is not about state imposition of criminal penalties. Yet this is another source of popular confusion. An act that violates the criminal law requires, in at least one setting, proof beyond a reasonable doubt before sanctions may be imposed. Possibly as a result of this, public debate often mistakenly assumes that it is the nature of the *act* that occasions a stringent standard of proof, rather than the context of the decision. Yet where the consequences of an adverse decision are other than deprivation of liberty or imposition of a criminal fine, there is no reason to suppose that the standard of proof should be the same even though the underlying act is identical. Thus, criminality under the law is not typically material to the epistemological question in nonjudicial decision making. We can question whether committing larceny is disqualifying for the position of quarterback, whether murderers should have their books published, or whether presidents should be impeached for illegally disposing of toxic wastes without doubting the wisdom of the criminal prohibitions.

Conversely, many acts that are not crimes may be material to an adverse decision. A bridge player who peeks at his opponent's hand, an

[3] An interesting question is whether doubts about materiality should raise the standard of proof. The evidence literature recognizes uncertainty with respect to the proof of a material fact and distinguishes relevance from materiality. But it ignores the question of uncertainty about materiality, because this falls within the domain of legal uncertainty (see Gary Lawson, Proving the Law, 86 Nw. U. L. Rev. 859 (1992)) rather than factual uncertainty. Yet since questions of materiality can be uncertain, it seems plausible to compound the probabilities and so to demand less confidence in the proof of a plainly material fact than we would for the proof of a fact of less certain materiality.

orthodox rabbi who eats pork chops, and a dinner guest who picks his nose at the table have committed no crimes. Yet these charges, if true, would be material to a decision to ask the bridge player to be a fourth, to dismiss the rabbi, or to invite the guest to a dinner party. Thus, our assumption of noncontroversial materiality depends neither on criminality nor on the possibility that the alleged conduct would be wrongful in any other context; materiality is a decision-relative concept. Doing a favor for a friend is often desirable, but not when it is a judge imposing an unusually small fine on a pal. Deliberate deception is usually wrong and sometimes unlawful, but a talent for deception would hardly be disqualifying were we seeking to employ a professional poker player.

We do not limit our inquiry to charges that, if true, would necessarily be dispositive. For example, although a charge of negligence in attending faculty meetings may be material to deciding whether to hire a senior faculty member, few would consider poor attendance dispositive. What one should clearly *take into account* is not necessarily what one should, all things considered, *do*.[4] So although the weight of some factor in the final decision might determine the standard of proof to be applied to determine that factor's existence,[5] we simplify the analysis by addressing

[4] The distinction between what is to count in a decision and what, on balance, to do is well-established in the literature on rights, duties, and obligations, since what one has a right, duty, or obligation to do is not necessarily what one should do, all things considered. See Joseph Raz, Practical Reasons and Norms 25–28 (1975); W. D. Ross, The Right and the Good 30–32 (1930); Frederick Schauer, Playing by the Rules: A Philosophical Examination of Rule-Based Decision-Making in Law and in Life 5–6, 113–15 (1991); John Searle, Prima Facie Obligations, in Practical Reasoning 81 (Joseph Raz ed. 1978); A. John Simmons, Moral Principles and Political Obligations 7–11 (1979). Again, we note the interesting complication that we might wish to demand greater confidence for factors that play a larger role in the ultimate decision. For example, suppose we hear from a peccable source that a candidate for a faculty position is lax about attending faculty meetings at her current institution. We might consider the weak evidence we have as sufficient just because this material fact would, if true, still play only a small part in the ultimate decision. Yet were this a candidate for dean, where the charge might be of much greater moment, perhaps we should demand greater confidence in the existence of the alleged facts. Still, this first reaction may be incorrect: Insofar as the degree of materiality is determined under conditions of certainty, lowering the probability threshold for low-materiality charges compared to the probability threshold for high-materiality charges could distort the previously determined materiality calculus by implicitly increasing the weight to be given to some factors. Suppose, for example, that faculty meeting attendance can only earn a candidate a maximum of five points out of 100, and responsibility in meeting teaching obligations counts for a maximum of 10 points. If a candidate loses the meeting attendance points on a .3 likelihood and the teaching responsibility points only on a .6 likelihood, then for cases at the respective minimum thresholds, the two factors achieve identical weight.

[5] When multiple factors are material to a single decision, applying the same standard of proof to each factor that is to be applied to the decision itself can cause the degree of confidence in the final decision to fall below the postulated threshold. Ronald J. Allen, A Reconceptualization of Civil Trials, 66 B.U. L. Rev. 401 (1986). Yet raising the standard

here the question of degree of confidence for all facts whose existence would be material to a decision, not just all facts whose existence would be dispositive.

Finally, we focus on the *how much* and not the *how* of determining the existence of some fact. Questions about the reliability of hearsay evidence, about so-called circumstantial evidence, about documentary evidence rather than oral testimony, and about sworn as opposed to unsworn testimony are not our focus. Under any method of obtaining and evaluating evidence, the sum total of that evidence (including greater quantities of less reliable evidence) can be cumulated into a degree of confidence in the truth of a factual proposition. Our concern is with this degree of confidence, not the methods that produce it.

III. Lowering the Standard of Proof

The degrees of confidence employed in the legal system—proof by a preponderance of the evidence, proof by clear and convincing evidence, and proof beyond a reasonable doubt—can be numerically quantified for analytic purposes.[6] We take the preponderance of the evidence standard as equivalent to a .51 probability, proof beyond a reasonable doubt as

---

of proof for each factor may also have the effect of raising the standard of proof for the decision itself. These complexities may make it impossible to imagine an ex ante specification of the burden of proof for individual factors; that burden of proof will vary with the actual degree of confidence the decision maker has in the other factors in the decision.

[6] Some would disagree. See, for example, David H. Kaye, Do We Need a Calculus of Weight to Understand Proof Beyond a Reasonable Doubt? 66 B.U. L. Rev. 657, 667–68 n.22 (1986). Here we merely announce the side of the existing debate to which we subscribe rather than go off on a detour to explain why. For more extensive treatments, see Richard Lempert, The New Evidence Scholarship: Analyzing the Process of Proof, 66 B.U. L. Rev. 439 (1986); Symposium, Decision and Inference in Litigation, 13 Cardozo L. Rev. 253 (1991).

We avoid the term "presumption of innocence" throughout this article. Although the presumption of innocence is typically associated with the criminal process and its requirement of proof beyond a reasonable doubt, the presumption of innocence is not a proof standard of all, being consistent with each of the different proof standards noted in the text. In few legal standards is the defendant initially required to establish nonliability or nonguilt, and thus the term "presumption of innocence" is as compatible with a system requiring only a preponderance of the evidence to overcome the presumption as with a system requiring proof beyond a reasonable doubt.

Similarly, the Scottish verdict of "not proven" in criminal cases does not appear to be a distinct standard of proof; rather, it is vehicle by which a jury that has determined that the prosecution has not proved its case beyond a reasonable doubt can express its view that the defendant more likely than not (a preponderance of the evidence) committed the acts charged. The unavailability of such a verdict in the United States seems related to the tendency of acquitted public figures to claim vindication by virtue of the "not guilty" verdict, a much more difficult claim after a "not proven" verdict.

roughly .95, and proof by clear and convincing evidence[7] even more roughly as perhaps .75.[8]

It is well understood that a legal system's choice among these standards is an exercise in trading off the harms that flow from different types of error.[9] Blackstone's maxim that "it is better that ten guilty persons escape, than that one innocent suffer"[10] reflects the greater harm ascribed to the conviction of the innocent than to the nonconviction of the guilty. Actual weights for different errors, however, vary with the prior probabilities. If a person actually charged is more likely than someone selected randomly from the population to be guilty of the crime, the calculus of errors will be different than if the prior probability of guilt for someone charged is no greater than that for a randomly selected citizen. Even so, as long as a wrongful deprivation of liberty is more harmful than nonpunishment of those who deserve it, any version of Blackstone's maxim[11] will generate something close to the standard of proof beyond a reasonable doubt. In contrast, the preponderance standard used in most civil litigation reflects the view that a failure to find for a deserving plaintiff is no less harmful than holding liable a nonculpable defendant.[12]

[7] This is the standard required for the civil commitment of the mentally ill and for proof of actual malice in public figure libel cases after N.Y. Times Co. v. Sullivan, 376 U.S. 254 (1964).

[8] On the relationship between numerical probabilities and the ordinary language of probabilistic assessment (words like "likely"), see Frederick Mosteller & Cleo Youtz, Quantifying Probabilistic Assessments, 5 Statistical Sci. 2 (1990).

[9] See, for example, John Kaplan, Decision Theory and the Factfinding Process, 20 Stan. L. Rev. 1075 (1968); David Kaye, The Limits of the Preponderance of the Evidence Standard: Justifiably Naked Statistical Evidence and Multiple Causation, 1982 Am. B. Found. Res. J. 487; Richard Lempert, Modeling Relevance, 75 Mich. L. Rev. 1021, 1032–34 (1977). There is a question, not important here, whether this variety of error analysis is sufficient, or if instead a better expected value analysis must weigh the expected benefits of correct decisions as well. Richard Friedman, Standards of Persuasion and the Distinction between Fact and Law, 86 Nw. U. L. Rev. 916, 927 n.28 (1992). Still, we believe that a properly conceived regret matrix, which measures comparative utility losses from different mistakes, fully captures the proper comparison between two decisions. *Id.* at 927–28 n.29.

[10] 4 W. Blackstone, Commentaries on the Laws of England *358.

[11] Others have used different ratios for the comparative utility of convicting the guilty and freeing the innocent. Fortescue and Paley suggested 20 to 1 (J. Fortescue, Commendation of Laws of England 45 (F. Grigor trans. 1917); W. Paley, Moral and Political Philosophy, in Works 27 (1983)), which is close to our .95 quantification of proof beyond a reasonable doubt. (To be precise, a .95 probability is the appropriate benchmark when the cost ratio for the two types of errors is 19 to 1, since $1 \times .95 = 19 \times .05$. We approximate with 20 to 1.) Hale chose 5 to 1 (2 M. Hale, Pleas of the Crown 289) and Starkie 99 to 1 (1 T. Starkie, Practical Treatise on the Law of Evidence 506 (4th Am. ed. 1832)).

[12] Things are a bit cloudier in practice. Because most causes of action have multiple elements, requiring a plaintiff to establish *each* by a preponderance of the evidence builds in a defendant bias. See Ronald J. Allen, The Nature of Juridical Proof, 13 Cardozo L. Rev. 373 (1991).

Proof beyond a reasonable doubt represents a high burden for the prosecution, so it can be seen as setting a low burden for the defendant. If .95 quantifies the reasonable doubt standard, a defendant can establish a reasonable doubt by proving to .06 confidence that he did not commit the act charged. Of course, in a criminal proceeding the defendant is literally not required to establish anything, in terms of bearing a burden of producing evidence; nonetheless, this allocation of the onus of coming forward does not detract from the relationship by which the prosecution's inability to establish guilt to a confidence level of .95 mirrors the defendant's ability to establish nonguilt to a confidence level of .06.

If a probability of innocence of .06 is sufficient to justify acquittal in the criminal context, it follows that where the disutility ratio is reversed, .06 is sufficient to convict. Suppose there is a reasonable possibility (a positive characterization of reasonable doubt)—.06—that an applicant for a teaching job is a sexual harasser of students. Suppose also that there are several applicants and that each can be rejected nonpublicly and without stating reasons. Under these circumstances, the harm from mistakenly hiring the sexual harasser would bear the same relationship to the harm of mistakenly not hiring a teacher who does not harass as the harm of convicting the innocent bears to the harm of acquitting the guilty. It follows that a .06 probability would be sufficient to justify a finding of "guilt": a nonhiring where there would otherwise have been a hiring.

In considering the use of such low probabilities, it is important to distinguish among "degrees of innocence," the extent of culpability short of complete guilt as charged. Often we act on low probabilities of guilt only when the probability that the subject did at least something culpable is much higher. Consider the "appearance of impropriety" standard often applied to the conduct of public officials. Sometimes we do sanction someone totally innocent for behaving in a manner that appeared to others, mistakenly, to suggest wrongdoing.[13] At other times, however, the appearance of impropriety standard is (misleadingly) employed when the degree of confidence in the defendant's guilt is less, but where there is no doubt that the defendant engaged in some questionable behavior. Thus, in many contexts we seem especially willing to rely on low probabilities

[13] This is especially true when the subject knows of the appearance of impropriety standard, because then we can say that knowingly or negligently creating a false appearance of impropriety is still an instance of bad judgment. The case of Judge Haynsworth may be an example. In 1971, his nomination to the Supreme Court foundered on charges that he had sat as a judge in cases in which he had a remote financial interest, much stronger evidence of bad judgment than of any influence on his decisions.

when there is a high probability that the subject did something wrongful, even if less wrongful.

Suppose our .06 degree of confidence that an applicant is a sexual harasser is based on a much higher degree of confidence in the applicant's having engaged in unusually frequent touching of students in a previous job; we may be comfortable applying the .06 standard because of the much higher probability that at least something is amiss. But what if our only information were that this applicant is named Ralph Bizzle, that a teacher named Ralph Bizzle was once convicted for child molestation, and that 17 teachers in the country are named Ralph Bizzle? Here the probability of guilt is still .06, but we might hesitate to act on it (assuming for the moment that we cannot conduct further investigation) because now there is nothing even faintly inappropriate embedded in the remaining .94. Where innocence is truly innocent, therefore, as when the question is identity, it may be appropriate to employ a higher burden of proof than where innocence is compatible with "not quite guilty," as in many cases that do not involve questions of identity but only how close the defendant's conduct was to the border between the culpable and the nonculpable.

A willingness to employ low-probability standards of guilt outside the legal process is straightforward decision theory and hardly revolutionary for the theory of proof. Consequently, the common unwillingness actually to use low probabilities may be based on an implicit disutility of convicting the innocent that is higher than most standard quantifications suppose. When someone who is erroneously found to have committed wrong x is still guilty of having committed wrong y, our regret is less, and therefore when we are confident that the person is guilty of some wrong, we may be more willing to act on the basis of low probabilities for some related wrong. Would we feel that a great injustice had been committed, for example, were we now to discover that Al Capone had not evaded payment of income taxes, the only crime for which he was convicted?

Other factors may also complicate the analysis. Consider first the distinction between decisions made openly and often collectively (the Thomas hearings, for example) and the externally invisible decisions we make every day. When we decide whom to invite to dinner, whom to befriend, and what businesses to patronize, we test factual hypotheses: some we determine probably to be true, others we determine probably to be false, and about others we remain genuinely uncertain. In such circumstances, the consequences of decisions adverse to individuals are often small precisely because there is no widespread awareness of the grounds for the decision, or even that a decision was made. When our suspicion that A cheats at golf leads us to invite B to make up a foursome,

the harm to A is typically quite small.[14] More precisely, the harm to A is a lost opportunity, but in this context few other harms are attached to or the consequence of the harm of the lost opportunity.

Many decisions, however, are more public than a noninvitation or a nonappointment. Sometimes decisions are public because the consequences are public (for example, someone gets fired), sometimes because the decision must be publicly explained ("Why is George no longer in your long-standing tennis game?"), and sometimes because the decision-making process is open. In any of these cases, and especially when these dimensions of publicity are combined, the publicness will highlight the charges of misconduct, which is less likely to occur when decisions are less public.[15] Moreover, public decision making will increase the spillover consequences of the immediate decision. Partly this is due to stigmatizing effects, discussed below, and partly to the limits on how complex categories of decision can be before they become unusable. Some theoretically sound distinctions might be difficult to draw in practice. Thus, when there is reason to believe that theoretically desirable decision-making procedures will be misused, they may become undesirable.[16] Say there

[14] We should, however, distinguish harms of insult from harms to reputation. See Frederick Schauer, The Phenomenology of Speech and Harm, 103 Ethics 635 (1993). If A expects to be invited and is not, there is a harm to A greater than the harm of not getting to play golf, even if the noninvitation is never communicated. But the harm to A coming from this insult is exacerbated when the noninvitation, and the reasons for it, are communicated to third parties in a way that damages A's reputation. Thus the noncommunicated insult lies between the case of noninsulting noncommunication (most of us have probably not been offered far more jobs than we know) and the case of both insulting and reputation-harming public denials.

[15] Consider how odd it seems to say that John Doe is so much better a philosopher than Jane Roe that he should be hired even though we are convinced he is a sexual harasser. And it would be harder still to say that John Doe is enough better of a philosopher that he should be hired even though there is a .4 chance that he is a sexual harasser but enough better of a philosopher that he should be hired were there a .8 chance that he is a sexual harasser. Although we believe this kind of implicit reasoning is common, it is often difficult to articulate such reasoning publicly. This might be because the quality of the reasoning is defective, or because misconduct like sexual harassment is dispositive, even though factors such as poor collegiality might not be. It might also be, however, that the fact of publicity is attention grabbing and makes public charges of misconduct more distinct from other factors in the same decision than they would be without the publicity. When charges are not public, however, and especially when decisions are not bipolar, things look different. How friendly should we be to A in light of the charges against her? Should we give more of our trade to B and less to C because we have some reason to believe that C does not honor guarantees scrupulously, even though we know that B's prices are higher than C's? With a spectrum of choices, the question of how much to weigh factor *f* is some combination of how strongly we believe *f* to be true and how important *f* is, if true. In other cases—the ones we focus on for ease of analysis and exposition—the epistemological question about one factor stands phenomenologically apart from other factors, a separation typically fostered by public decision making.

[16] One manifestation of this is the slippery slope phenomenon. See Douglas Walton, Slippery Slope Arguments (1992); Frederick Schauer, Slippery Slopes, 99 Harv. L. Rev. 361 (1985).

is a fear that many people already make decisions on the basis of too-low probabilities of wrongdoing (as with insufficiently grounded rumors). Then the public use of very low degrees of confidence may entail unacceptable risks that appropriate low probabilities will signal to others the permissibility of employing what are in fact (but not to them) inappropriately low probabilities. If so, there might be a lower limit of probability for a public determination of misconduct, even though it would be artificial and mistaken without taking account of the possibility of mistaken application.

Stigma is the other major consequence of publicity. When decisions are nonpublic, the consequences of a finding of wrongdoing are typically limited to that decision. For example, the teacher we do not employ may find a job elsewhere. However, the consequence of an adverse decision publicly announced to be based on wrongdoing can be greater; the reasons for the decision may create a stigma that contaminates many decisions about the same individual.[17] One ground for the argument that the burden of proof ought to have been quite high in the Thomas hearings is that being publicly found to have been a sexual harasser has stigmatizing consequences well beyond being denied the position of Supreme Court Justice, and well beyond the negative consequences of having been found to have unpopular views about constitutional interpretation (Bork) or to having used marijuana as an adult (Ginsburg). If so, then the expected harm to the individual of a mistaken determination of guilt becomes greater, though the expected harm to the public of a mistaken determination of innocence remains unchanged.[18] As a consequence, the degree

[17] This is especially true when having been considered but rejected would otherwise have put the subject in a better position than she was in before having been considered. Being known to have been in the final three for an academic appointment, even if one is not appointed, will help the reputation of someone not previously thought of as being in the top three, even though it would harm someone who was previously thought of as the best in the field. Yet publicity about the reasons for nonappointment will typically reduce the positive effects for the former and heighten the negative effects for the latter, especially if those reasons pertain to misconduct.

[18] It appears more injurious to the public to have a sexual harasser as a Supreme Court Justice than to have a former marijuana user as a Supreme Court Justice, and more injurious to the individual to be found to have been a sexual harasser than to have used marijuana. In other cases, however, the degree of harm to the public from having a person with characteristic c in office may not parallel the harm to the individual of being found to have characteristic c. Consider charges that one nominee for a seat on the Supreme Court had engaged in plagiarism in law school and as a law professor and that another nominee embezzled money from the school newspaper. Putting aside the symbolic legitimation effects (which would be roughly the same in the two cases), it is plausible to suppose that the negative reputation effects of the charges are approximately the same in the two cases, but that the harm to the public of having a plagiarizer as a Supreme Court Justice is greater than the harm to the public of having an embezzler as a Supreme Court Justice.

of confidence necessary to justify an adverse decision must be raised. Conversely, when stigmatizing concerns are minimal (as perhaps with charges made to a president considering whom to nominate), this reason for elevating the burden of proof is eliminated.

Public decisions are also likely to increase reliance on the status quo in determining the appropriate burden of proof. While status quo bias[19] is not always appropriate, reliance and stability values will often suggest a preference for that which exists just because it exists, and such a preference for the status quo elevates the degree of confidence required to change present arrangements. Thus, the burden of proof should be higher for firing an employee for misconduct than for refusing to hire an applicant based on the same charges. Sometimes these status quo considerations are based on systemic values of stability and sometimes on the affected individual's interest in reliance and repose, an interest often translated into the terms *entitlement* or *vesting*—a property right in the status quo. If it is more difficult for people to readjust various tangible aspects of their lives than to compensate for the disappointment of an initial denial, their current situation will secure them an element of entitlement. Consequently, the degree of confidence in suspected misconduct will be greater to remove this entitlement than to refuse to enter into the relationship at the outset.[20]

Although a status quo bias thus generates epistemic conclusions, it may also be based on epistemic factors. Assume that committing a disqualifying act reflects a propensity to commit such acts and that the probability that any disqualifying act will be discovered and reported is greater than zero. Then the fact that no such acts have been reported during a relationship is some evidence of nonpropensity, the effect of which is to generate a bias in favor of continuing the relationship.[21] More-

[19] See William Samuelson & Richard Zeckhauser, Status Quo Bias in Decisionmaking, 1 J. Risk & Uncertainty 7 (1988).

[20] The timing of the charges against Thomas might have produced an entitlement effect. Because Hill's charges became public after considerable public discussion of Thomas's qualifications, some senators and citizens who had already concluded that Thomas should be confirmed (however technically nonfinal that conclusion was) might have seen that conclusion as sufficient to elevate the burden of proof. This may illustrate that decisions are themselves reasons for action, thus creating presumptions in their favor. See Joseph Raz, Practical Reasons and Norms 65–73 (1975); Schauer, *supra* note 4, at 110–14.

[21] There could, of course, be a history, whether of misconduct or of good conduct, at some other institution. This raises the question whether most people feel that it is worse to sin against them than to sin against another. When there is the same confidence that student A plagiarized elsewhere as there is that student B plagiarized at our own institution, we may take B's act more seriously than A's. This may stem from doubting the epistemic equivalence of the two acts; we may have more confidence in our own investigations than in those of others. Apart from this, however, it would seem to make more sense to treat the foreign misconduct more seriously than the domestic, for comity considerations would

over, if we consider as qualifications positive performance characteristics as well as the nondisplay of negative performance characteristics, then generally the evidence of ability to perform the job is greater for the person who is actually performing it than for the person who is performing a job at least slightly different at another location or institution. Finally, retaining an employee keeps open the option of future dismissal, but dismissal ordinarily precludes rehiring.

While stigmatization and entitlement effects raise the costs of a mistaken finding of guilt, the costs of mistaken acceptance into some position (child molesters as teachers, pilferers as shop attendants, plagiarizers as students) are analogous to the costs of unimprisoned criminals. These costs may be greater when we take into account the endorsement effect; holding a public position is seen as public approval of the conduct of the holder (even if it was not previously known). Thus the costs of mistaken acceptance are the costs of seeming validation of wrongful acts, the costs of nonretribution of acts that deserve retribution, and the costs to institutions or potential victims when misperforming employees continue to serve.

At times the costs of misconduct are not concentrated on particular victims (such as the victims of child molesters or sexual harassers) but are distributed substantially more widely across a group. In such cases, the costs may well be undervalued in a disciplinary process. A rude remark to 20 people may be more serious than a clear offense to one, but with free-rider problems, and the likelihood that the remark falls below some predetermined threshold, the remark may escape censure.

Actions and inactions are often influenced by a trade-off between magnitudes and numbers. If there is a dangerous point in the road that would cost $200,000 to repair, and 100 mildly dangerous points that would cost $1,000 each to repair, and if the expected harms from the one very dangerous point are the same as from the hundred mildly dangerous points, there is some evidence that people will choose to repair the one point over the one hundred.[22] Similarly, people might prefer freeing 100 individuals, each of whom has a .01 probability of committing a murder within the next year, to freeing one person who has a .50 probability of committing a murder within the next year, even though there are twice as many expected murders if 100 people are freed. Such a tendency to undervalue the costs of mistaken vindication in cases of broad distribution, however,

---

suggest a reluctance to reevaluate decisions made by others, reluctance typically not present with respect to reevaluating our own decisions.

[22] Amos Tversky & Daniel Kahneman, The Framing of Decision and the Psychology of Choice, 211 Sci. 453 (1981). See also Charles Fried, An Anatomy of Values (1970).

might be counterbalanced by an offsetting tendency to undervalue the costs of a mistaken finding of wrongdoing in cases of broad distribution. Because most people feel greater guilt when they wrongfully punish someone than when they are part of a society that wrongfully punishes someone, a political process that secures comfort in numbers and undervalues the cost of mistaken acquittal is likely to undervalue the cost of mistaken conviction. The actual outcome will depend on the balance of these two factors.

If stigmatization and entitlement effects do not play a role, and if framing effects are not significant, then to establish the degree of confidence needed to take an adverse decision, we employ a straightforward assessment of the expected harms of different types of error. In the polar case, therefore, one in which an *applicant* for a position involving great responsibility to third parties is charged nonpublicly with having committed acts that directly relate to the responsibilities of the position, and if the adverse decision would also be nonpublic, then perhaps no more than a slight possibility—say, .06—that the charges were true would be sufficient for making an adverse decision. Ideally, in such cases we would engage in further investigation, and it is plausible to suppose that the possibility of acting on low probabilities would impose an increased obligation for further investigation before so acting. Still, investigation is costly, and even further investigation may still produce low confidence levels. Yet just as there are nondisciplinary contexts in which we comfortably rely on low probabilities—as when we delay a flight because of a low probability that there is a bomb on board the airplane—there may also therefore be contexts in which no more than a slight possibility would be sufficient to support a finding of individual wrongdoing.

IV. The Ethics and Mathematics of Compound Probabilities

Whatever degree of confidence is required to justify an adverse decision, there will always be cases in which there is insufficient confidence based on a single observation or single charge. In the criminal law, this situation properly produces an acquittal. Yet often, both within and without the criminal justice system, there are multiple charges of similar but distinct acts, with each charge producing its own decision. Suppose we conclude that some class of employees, say school teachers, should be dismissed if there is a *significant possibility*—say, .35—that they have committed an act of sexual harassment. And suppose that a charge is made against a particular teacher, a charge the decision maker believes is true to a probability of .1. Finally, assume that the only options are to either dismiss the teacher or the charge. With these as the only options, the charge will be dismissed.

But suppose now that two other charges of equal magnitude are made, that the three are independent of each other,[23] and that a dismissal of one has no effect on the resolution of any of the others. If these three events were truly independent, as they would be, say, if they were charges by three people who are unaware of the charges by the others, and if there were no tendency of particular individuals to engage in sexual harassment, the probability that at least one of the charges is true is .271 ($1 - (.9 \times .9 \times .9)$), but because each of the decisions is independent, the teacher could not be dismissed. This holds true even with seven independent charges, producing no dismissal even with a more likely than not (.52) probability ($1 - (.9 \times .9 \times .9 \times .9 \times .9 \times .9 \times .9)$) that the teacher has committed an act of sexual harassment.

This is a very crude example; in reality, some people are more likely to harass than others, so one would be more likely to believe a charge if someone had been charged in the past. Let us assume that before learning anything we thought the odds that someone was a harasser were 1 to 99 against. Assume also that a harasser is, continuing the above example, 11 times more likely to be charged than is an innocent party. With one charge, therefore, the odds that the teacher is a harasser are $11/1 \times 1/99$, or 11 to 99 against. This means that on average in 11 cases out of 110 (11 + 99), a singly-charged person will be a harasser, which (as expected) is consistent with the .1 probability in our example. When second and third charges are received, the odds multiply substantially: an individual charged three times has an odds ratio of 1,331/99, which translates into a .931 probability that he is a harasser.

But we believe that individuals differ substantially in their likelihood of engaging in guilty behavior. If so, the truth of accusations is no longer independent. So we ask a new question: How likely is the individual to be a harasser? (Previously, we asked whether he had committed a specific act of harassment.) We seek to compute a terminal probability, based on the evidence, that the individual possesses the guilty trait.

Note that even if the individual has the trait, all of the accusations may be false. A medical analogy may be helpful. Chest pains are indicative

[23] The assumption of independence is complex, even though we treat the probabilities of individual events as independent. Suppose that on 5 different days there is a shortage in the cash drawer of a particular shop employee, and suppose we think, based on general knowledge about such matters, that a shortage indicates employee skimming to a probability of .5. But then suppose we monitor the employee 1 day and discover a shortage but no skimming. This discovery would not leave the other .5's in place, but would suggest that all of the .5's must be readjusted downward. If a probability affecting multiple independent events is recalibrated, it will appear as if the events are in fact dependent. See John W. Pratt & Richard J. Zeckhauser, Inferences from Alarming Events, 1 J. Pol'y Analysis & Mgmt. 371 (1982).

of heart disease. An individual with heart disease may suffer multiple chest pains, but the pains may not have arisen from heart disease. If this were a criminal case, the possibility of drawing a correct conclusion based on false accusations would disturb us greatly. For most of the potential adverse decisions considered here, however, that concern is lower. Not hiring an embezzler whose past charges of embezzlement were false would not seriously disturb most bank presidents.

To reiterate, our goal is to deal, outside of the criminal justice system, with multiple charges, each of which is insufficient to justify a finding of guilt. We assume, as is often the case, that the guilty behavior persists throughout the interval of observation—that an individual does not change his behavior once accused, and that there is no prior history.[24] The analysis then begins with a prior probability of guilt, p, for an individual about whom we have no information. To simplify the analysis, we divide the world into individual units. For sexual harassment, the units might be thought of as individual occasions or encounters; for other behaviors, a specified time period might be more appropriate. We know only whether an individual is accused within a unit. (One or multiple accusations within a unit are treated the same.) The likelihood that a guilty party will be accused within a unit is A_g. The likelihood that an innocent party will be accused within a unit is A_n. Thus, the likelihood that a guilty party will *not* be accused within a unit is $1 - A_g$, and the likelihood that an innocent party will not be accused within a unit is $1 - A_n$.

Now, let $p(m, n)$ be the probability that an individual who is accused in m units out of n is actually guilty.[25] If we have the experience of only one unit, and the individual is accused during that unit, the probability of guilt is computed by Bayes' Theorem as follows:

$$p(1, 1) = pA_g/(pA_g + (1 - p)A_n).$$

Here the numerator gives the likelihood of an accusation and guilt; the

[24] We use the term "accusation" in the sexual harassment context. Were the charge one of employee pilfering or student plagiarism, we might refer instead to a potentially incriminating observation.

[25] The formula for the terminal probability given m accusations in n units is easily derived using the binomial formula and Bayes' Theorem. If the probability of an accusation in a unit is A, the probability of getting m accusations in n units, by the binomial formula, is

$$B(m, n/A) = C(m, n)A^m(1 - A)^n,$$

where $C(m, n)$ is the number of combinations taking m things n at a time.

The terminal probability of guilt given m accusations in n units is, by Bayes' Theorem,

$$p(m, n) = (pB(m, n/A_g))/(pB(m, n/A_g) + (1 - p)B(m, n/A_g)).$$

denominator gives the likelihood of an accusation, whether the individual is guilty or not guilty.

Simplify, and consider a situation with m periods, each with an accusation. Following the same logic, this yields

$$p(m, m) = pA_g^m/(pA_g^m + (1 - p)A_n^m).$$

Let us assume that the prior probability of guilt is .02 and that $A_g = 10A_n$; that is, a guilty party is 10 times more likely to be accused in a period than is a nonguilty party. Plugging in the values given, if there is 1 period of accusation, $p(1, 1) = .17$. If there are three charges in each of 3 periods, then the terminal probability of guilt, $p(3, 3)$, rises to .95. Thus, the confluence of even a few events which would be inconclusive were they to occur just once may together comprise overwhelming evidence of guilt. More complex calculations can be undertaken for cases in which there are some periods without accusations or if more than one accusation in a period is counted differently from one accusation in a period.[26]

The lesson of the computations, therefore, is that a teacher who is assessed separately on three charges, and escapes action because the terminal probability for each charge is only .17, is still .95 likely to have committed an act of sexual harassment. Thus it may seem reasonable to suppose, for example, that a school principal, most likely proceeding more informally and less mathematically, would significantly increase the terminal probability in response to each additional charge. Raising the terminal probabilities in this fashion would appear appropriate even if the evidence in each of the successive cases were equivalent.

A critical simplifying assumption in this analysis is that the accusations are conditionally independent.[27] In many instances, however, there is a positive dependence among accusations, for both guilty and nonguilty parties, based on the assumption that accusations tend to beget further accusations. Where there is such a positive dependence, the terminal probabilities will escalate less rapidly. Still, the point remains that probabilities that would be insufficient when taken alone may well produce quite high probabilities of guilt when cumulated.

In light of this, it might seem unrealistic to suppose that a decision maker would ignore the evidence from one charge in assessing the proba-

[26] Only the relative values of the accusation probabilities matter for the calculations shown here. Were there periods without charges, absolute values would be required to make the appropriate calculation using the binomial formula.

[27] Conditional independence means that given the accused is of one type, say, not a harasser, the likelihood of an accusation in each unit is independent.

bility of another. Yet much real-world decision making separates accusations and decisions much more than it cumulates them. The most obvious example, of course, is the judicial system. Although there are some circumstances in which prior similar acts are admitted into evidence as a factor in the determination of guilt or innocence,[28] the general "propensity" rule is that "[e]vidence of other crimes, wrongs, or acts is not admissible to prove the character of a person in order to show action in conformity therewith."[29] Thus, charges are typically evaluated independently of previous charges, and previous convictions (and certainly acquittals, even though the fact of the charge, and the fact of nondismissal at an early stage, will substantially raise the prior probability of guilt) are inadmissible in a trial for subsequent charges. Accordingly, no sanction is imposed even if five consecutive juries in five different cases decide that the defendant has not been proved beyond a reasonable doubt to be guilty. This is so even if each jury has determined that there is a .94 probability of guilt, creating an extraordinarily high probability, verging on certainty, that the defendant has committed at least one of the acts charged.

Of course, the practice of noncumulation of charges in the criminal law

[28] One example is so-called signature crimes. If the defendant has on previous occasions been known to commit bank robberies carrying a Neptune's trident as a weapon, wearing a fish mask and clothes of artificial scales, then if such characteristics were present for the crime charged, the previous events establishing the pattern of behavior are admissible into evidence. Similarly, the "doctrine of chances" (see R. v. Smith, 11 Cr. App. R. 229, 84 L.J.K.B. 2158 (1915)) admits prior similar acts when those acts are a sufficient "earmark" that they strongly rebut the possibility of absence or mistake, as when a man's three wives all drowned. In other circumstances prior acts may be admitted to show "motive, opportunity, intent, preparation, plan, knowledge, [or] identity." F.R. Ev. 404 (b). Rules 413–15, added by the 1994 Criminal Justice Act, expand the admissibility of prior acts of sexual misconduct. As we write this, public attention is highly focused on the trial of O. J. Simpson, and various questions about the admissibility of prior acts have been raised in the trial. This is interesting not only as a public display of the operation of the rules of evidence but also as a vehicle for influencing public views about the propriety of admitting past acts as proof of current misconduct or as predictors of future misconduct. Insofar as public perceptions about the usability of past acts are influenced by this trial, or other prominent trials, this influence would support our central thesis about the potentially distorting effect of perceptions about criminal trials on the design of institutions, outside of the criminal justice system, for determining the existence of wrongdoing.

[29] F.R. Ev. 404(b). One reason for the rule is that evidence of other crimes is thought to undermine the presumption of innocence. See Government of Virgin Islands v. Toto, 529 F.2d 278, 283 (3d Cir. 1976); 2 McCormick on Evidence, at sec. 190 (4th Ed. John W. Strong ed. 1976). But the rule also applies in civil cases, Hirst v. Gertzner, 626 F.2d 1252, 1262 (9th Cir. 1982); 2 McCormick on Evidence, at sec. 189, in part because of a belief that both in civil and in criminal cases the "dangers of prejudice, confusion and time-consumption outweigh the probative value." Boyd v. United States, 142 U.S. 450 (1892). See also Reyes v. Missouri Pacific Ry. Co., 589 F.2d 791 (5th Cir. 1979); Miller v. Poretsky, 595 F.2d 780, 783–85 (D.C. Cir. 1978).

serves important goals. If juries are prone to overassess the probative value of prior similar acts, or if there are independent moral and social goals served by "fresh start" opportunities for rehabilitation, there may remain good reasons for the criminal law's exclusion of probabilistically relevant evidence about prior acts. Obviously there are costs associated with these goals (which the foregoing statistical analysis may illuminate), but weighing the costs and benefits of the refusal to cumulate in the criminal process is not our goal. Rather, we argue that the benefits of noncumulation are less compelling outside the judicial context, a context that arguably includes civil as well as criminal trials. Often, for example, the consequences of negative decisions outside of court are less grave. Denial of serious consideration for a job is not the same as denial of personal liberty by imprisonment.

Civil proceedings aside, it is likely the particular salience of criminal trials in this society, and the accompanying concern for the rights of accused parties, that often leads to a presumption in nonjudicial contexts that decisions should not be cumulated and probabilities should not be compounded. Outside of the judicial context, however, the presumption of noncumulation is frequently far less justifiable. Consider those bureaucratic settings where, as in the judicial system, the same individual may be subject to serial evaluations by different evaluators in different settings. Suppose an employee is charged with sexual harassment and the employee's supervisor believes that the charge must be proved by a preponderance of the evidence (.51) before a negative decision will be taken. The supervisor investigates, concludes that the charges have (to use the same mathematics as used above) a .17 probability of being true, and a .83 probability of being false, and so takes no disciplinary action. Some months later, the same employee, who has been transferred, is accused of similar acts,[30] with the same assumptions, the same conclusions, and thus the same outcome reached by a different supervisor. A while later, after a subsequent transfer of the employee (or his supervisor), a third charge is made, again with the same assumptions and the same result. The consequence is that there now exists within the organization an employee .95 likely to have committed an act of sexual harassment, but as to whom no disciplinary action has been taken.

Is there any justification for treating differently an employee .95 likely to have committed a specific act of sexual harassment and another employee .95 likely to have committed at least one of three specific and identified acts of sexual harassment, even if we cannot in the latter case

[30] Assume the chargers and the charges are independent.

identify with confidence one of the three acts as having almost certainly occurred? Sometimes the answer may well be in the affirmative. For example, with respect to acts whose wrongfulness consists in negligence rather than wrongful intent, there may be good reasons not to aim for a zero negligence rate, for that may induce excess risk aversion and consequent suboptimal performance. More typical, however, are behaviors that are unacceptable even if they happen just once. There is a difference between the professor who teaches one class poorly and the professor who engages in one act of sexual harassment. Where a single act is disqualifying, or at least sufficient to justify some sanction, some of the special reasons for decision separation in the criminal law are inapplicable, and there seems little reason to believe that a probability sufficient to conclude that this disqualifying act has occurred should not also produce the same conclusion, and the same consequences, if phrased in terms of whether one disqualifying act has occurred, even if we are not as confident as in other circumstances that it was *this* act.

V. An Application

Although many judicial and nonjudicial contexts are poorly designed to take account of compound probabilities of events whose individual probabilities would be insufficient to justify adverse action, some social systems do recognize the importance of compound probabilities. At its best, perhaps, in the social system we call "gossip," low-probability charges are first merely remembered, but when similar charges are heard several times, the recipient cumulates the probabilities, informally, when deciding whether to pass along the charges to someone else[31] and what assessment to make when the charges are passed on.

[31] Consider the charges against former Senator Brock Adams, accused of sexual harassment by eight different women, quite possibly under circumstances in which the charges were independent. For the editors of the newspaper that learned of the charges, the unwillingness of the accusers to have their names disclosed might have reduced the probability of truth for any one charge below the level at which the newspaper would have felt comfortable printing the charges. However, the newspaper's eventual decision to publish may have reflected the extreme unlikelihood that all of the charges were false, assuming at least some independence among the charges.

There is an ambiguity here about what it is for a charge to be "low probability." If the prior probability, as perceived by the editors, that Adams had committed an act of sexual harassment was .01, then even if a charge were 10 times as likely to have been made if true than if untrue, the terminal probability after one charge is only approximately .1. But once the second charge is made, the prior probability is no longer .01 but .1, and a charge 10 times as likely to be made if true than if untrue, when applied to a prior probability of .1, produces a probability of approximately .5. Thus when we consider some charge to be a low-probability charge, it is usually because we assume a low prior probability. When such a low prior probability can no longer be assumed, retaining that assumption will yield inaccurately low terminal probabilities.

Somewhat more formally, the world of competitive contract bridge employs a "recorder" system during national championships. In that system, charges of improper signaling or other improprieties (such as taking information from a partner's hesitations), sometimes well-grounded but often not, are made to a designated recorder. The recorder can encourage the complainant to take the charge to a disciplinary committee, in which case the full range of due process procedures are applicable, or the recorder can simply informally note the existence of the first charge. When a second charge is made against the same player or partnership, the recorder takes account of the first charge in deciding whether to encourage the second complainant to bring the matter to the disciplinary committee. The decision to encourage is likely to reflect the recorder's cumulation of the probabilities of the first and the second charges.[32]

Still more formally, consider the sentencing practices of most U.S. courts. Of course prior convictions are relevant, but so are prior charges and prior acquittals. Given the safeguards provided by the use of grand juries and the rarity with which those totally innocent are prosecuted (Would you entrust your money to someone acquitted of embezzlement?), even an acquittal suggests a much higher than random probability that the defendant has committed the crime for which he was acquitted. As a result, the practice of considering previous charges and acquittals (with the presumed exclusion of instances of mistaken identity) at the sentencing stage can be seen as a way of considering compound probabilities.

With these examples as guides, consider again the typical workplace sexual harassment scenario. Given the unwitnessed and behind-closed-doors nature of most acts of sexual harassment, the likelihood that any given charge of sexual harassment will satisfy a decision maker to a high degree of confidence is likely to be lower than for acts more likely to be committed publicly or to have produced physical evidence. In fact, there is a whole category of wrongful acts—in addition to sexual harassment, other examples would be child molestation, spousal abuse, embezzlement, and plagiarism—that are likely to be committed multiple times and in which the charge of committing the act is likely to be establishable only to a low degree of confidence for a decision maker.

Given such patterns, we can expect that institutions treating charges such as sexual harassment as distinct events will not only be likely to

[32] In these cases, the disciplinary committee is not permitted to inquire into the first case, even though the existence of the first case was relevant to the decision by the recorder to encourage the complainant to bring the second case forward. But there is no a priori reason to suppose that earlier cases must be excluded from consideration in all settings where a similar system might be employed.

have some number of undisciplined offenders but will also be likely to have a higher proportion of undisciplined offenders than of undisciplined committers of other offenses for which charges are presented and resisted that lack the special epistemological features of sexual harassment and the other offenses noted above. When there are significant costs of reporting the offense, as with sexual harassment, the problem may be further exacerbated.

One remedy might be a recorder system different from the system used in competitive contract bridge. In our system, a designated official who is intentionally isolated from normal personnel processes (including both discipline and routine evaluation for promotion and reemployment) would receive, investigate, and record complaints. If the recorder determines that the complaint was totally unfounded, no action at all would be taken against the target of the complaint. If the recorder determines that the complaint is more likely true than not, it would be forwarded to formal disciplinary systems. But if the complaint were "possibly true," but not likely enough true to forward to the disciplinary mechanism, it would be recorded, unavailable to anyone unless further charges engendered sufficient confidence (with appropriate discounts for the possibility of nonindependence) to be recorded as well. If the number of cumulative charges reached a point (whether determined mechanically or entrusted to the recorder's discretion)[33] where the recorder was satisfied, by at least a preponderance of the evidence but possibly higher, that at least one incident had occurred, the record would then be forwarded for more formal disciplinary proceedings.[34]

[33] A mechanical determination may of course err, but that is the price of all rules, and one that is sometimes worth paying. The *Journal*'s referee asks us, rhetorically, to consider the possibility of a rule "that if there are x puffs of smoke there must be a fire." This seems less implausible to us than to the referee. As a decision rule, it might at times produce false positives and false negatives, but it still might be a good decision rule. So too with the recorder. A flat rule about moving to a further investigatory step after a specified number of charges might at times move unjustifiably to that next step and might at times not move to that step when it would have been correct to do so. But whether these mistakes of underinclusion and overinclusion are greater or lesser than the mistakes likely to be made with greater discretion is a genuine question, for one of the primary arguments for rule-based decision making in any domain is that the mistakes consequent on faithful application of blunt rules may at times be fewer than the mistakes made by unfettered decision makers within the area of their discretion.

[34] Considerations of due process (Was there an opportunity to defend against earlier charges? Is the act of recording one that itself triggers due process protection?) might affect whether the recorder would forward the full record or only the last charge. The possibility of due process before recording itself raises the even larger question of the process due in a wide variety of settings outside of the criminal trial. Although we bracket that issue here, it is worth noting that an increase in process at the recording stage will increase the availability of the evidence adduced at that stage for use in subsequent proceedings if further charges are made. But that same increase in process may also inhibit potential complainants from

Obviously many questions of implementation remain, many of them implicating important issues of procedural fairness that we do not address here. Would there be due process at the recording stage? Would previous recordings be available in formal disciplinary proceedings? Could recorded charges be the basis for lesser sanctions such as warnings?[35] Would such a procedure be likely to increase the willingness of justified complainants to complain and thus generate both more information and more unjustified complaints?[36]

These and other concerns are important, but we do not address them here because our primary goal is only to show that systemic accommodations of compound probabilities are far from impossible. Such accommodations tend to be rare in the process of determining guilt or liability in courts of law; for many good reasons societies suppose that their courts are places where acts and not persons are judged. Perhaps more accurately, courts are usually concerned with *tokens,* but in other contexts we are often concerned with *types.*[37] When we move out of the courtroom, however, and especially when employment is at issue—as with sexual harassment, padding of expense accounts, taking kickbacks, and pilfering of inventory—we are more concerned with evaluating persons than with focusing on individual acts.[38] This might be simplified as a concern with *character,* but that "characterization" may be oversimplifying. For when

---

coming forward with their charges, and might thus frustrate the goal of encouraging low-probability charges (as perceived by the third party evaluator of the charges) to be made available for possible cumulation.

[35] In some settings warnings are problematic, since a warning system puts people on notice that they will get one warning prior to the imposition of more severe sanctions. In some situations, such as student cheating, there may be good reasons not to create a system in which the optimal strategy is to cheat once.

[36] A recorder system may serve as a partial solution to the coordination problem pursuant to which complainants are unwilling to come forward unless they know other complainants are coming forward as well.

[37] Indeed, this distinction may explain some of the legal system's reluctance to admit so-called naked statistical evidence. In Cohen's Paradox of the Gatecrasher (L. Jonathan Cohen, The Probable and the Provable 73–76 (1977)), for example, our reluctance to impose liability on someone .501 likely to have been a gatecrasher may be based, in part, on a much lower probability of that person's having committed a particular instance of gatecrashing. Thus, when concerns are expressed about proportionate enterprise liability in cases such as Sindell v. Abbott Laboratories, 26 Cal. 3d 588, 163 Cal. Rptr. 132, 607 P.2d 124, cert. denied, 449 U.S. 912 (1980), part of the concern is about whether it is important that an individual specify on which exact occasion she was injured, and specify the particular act of the defendant's wrongdoing. However, our analysis and prescriptions do not depend on the desirability of their application to judicial contexts.

[38] The suggestion that the judicial/nonjudicial distinction maps neatly onto the token/type distinction is, of course, an oversimplification. Custody determinations (where low-probability charges may—and should—make a difference) are a good example of a judicial proceeding largely concerned with persons rather than with acts.

we are concerned with character, we are not usually concerned in Kantian fashion with the intrinsic worth of either a person or her acts. Rather, we use "character" as the simplified expression for those traits that lead to a propensity by certain people to commit certain kinds of acts. When we consider whether to admit evidence of character, therefore, we only consider whether to admit evidence of traits that are probabilistically related to the commission to certain acts. As we have stressed, however, the skepticism about the admissibility of those traits (or of previous acts that are evidence of the trait) that arises in judicial context does not always carry over to nonjudicial contexts as easily as is often supposed. One reason for this is that we are more often concerned with performance over time outside of the legal system than within it. When that concern dominates, two of the assumptions central to the legal system—that there is no liability on moderate or low probabilities of culpability, and that there is no compounding of low-probability charges—are substantially less compelling. In such cases, procedures very different from those employed in the courtroom are more justifiable. That is not because out-of-court determination of the truth of charges of misconduct should be less fair; it is because the notion of procedural fairness is itself domain- and decision-dependent. Where the decision-making goals differ from the goals of the system for determining criminal guilt or civil liability, the procedures of choice may differ as well. Indeed, although our discussion of cumulation has focused on the cumulation of low-probability charges, our first concern, low-probability charges, and our second, cumulation, are independent. The second is applicable to high-probability charges as well, though its bite is greatest in the low-probability scenario.

We have used sexual harassment as our primary example for a number of reasons. First, sexual harassment charges against prominent officials have recently attracted much attention; the charges against Senator Packwood have just been considered by the Senate, and charges against President Clinton are pending not only in a federal court in Arkansas, where they have been stayed, but in the public arena. Second, sexual harassment seems to us to be poorly dealt with as an act of professional misconduct. Third, the circumstances of most acts of sexual harassment are such that there is usually a high degree of uncertainty, so that the concerns that occupy us here are especially likely to be present. Some may disagree with us about the aptness of sexual harassment to much of what we maintain here. But none of our larger conclusions turns on the extent to which sexual harassment provides an appropriate example. As long as there are circumstances in which the noncompounding of highly uncertain conclusions produces suboptimal outcomes, there remain opportunities for substantial improvement.

VI. Conclusion

Many people invoke the metaphors, standards, and procedures of criminal trials when they have information relevant to potential adverse action against an individual. The information may be based on rumor, or circumstantial, or unsubstantiated, or based on the testimony of only one person. Although we believe that the standard of proof beyond a reasonable doubt was selected for criminal trials for sound reasons, we also believe that many of those reasons are less compelling or less applicable in situations such as dishonesty in the workplace, carelessness at a construction site, cheating at bridge, or plagiarism in the academy. Then different values are at work, and different harms stem from both wrongful conviction and wrongful vindication. We have argued, initially, that the idea of a "presumption of innocence" is so capacious as to be unhelpful and that degrees of confidence below, at times well below, that of proof beyond a reasonable doubt are ordinarily appropriate.

More significantly, in many situations individuals have repeated opportunities to engage in the same wrongful behavior. A careless supervisor may produce multiple accidents. A sexual harasser or child molester may prey on many victims. A plagiarizer may make plagiarism a habit. Although for good reasons, the system of criminal evidence is ill-equipped to evaluate people who may have committed multiple low-probability acts. We have argued for the soundness of decision-making systems that allow cumulation and have argued for evaluating such systems on their own terms and in terms of their own goals, without the distractions of the potentially distorting metaphors, slogans, and processes of the criminal law.

Although we deal only superficially with defining precise mechanisms for aggregating information across multiple wrongs, the central point is that there are numerous domains in which information other than that about the most recent accusation is appropriately brought to the decision maker's attention. The recorder system we sketch transcends the obstacle of compound probabilities and also encourages individuals to come forward with adverse information. This has advantages not only for determining the existence of past wrongdoing but also in the development of standards for the future. When the very definition of inappropriate behavior remains unclear, something like a recorder system may help a community define what it considers unacceptable and put on notice both those whose behavior is close to the line and those who would accuse unjustly. Even when final charges are not brought, a recorder system may provide a needed sounding board, defusing feelings of helplessness and forestalling the feeling that no one cares. Further exploration of these broader goals, however, is best left for another occasion.