# ENHANCING AND RESTRAINING ACCURACY IN ADJUDICATION

MICHAEL J. SAKS*

## I

## INTRODUCTION

The offerings of applied empirical social psychological researchers to the law tend to resemble those of civil engineers inventing new techniques for construction companies or electrical engineers adapting basic knowledge for the consumer electronics industry. They assume that their client institutions want more accurate measures, more robust materials, higher signal-to-noise ratios, more hits, and fewer misses. If the law is a decisionmaking machine, runs this implicit reasoning, then surely it will be eager to adopt findings and techniques that will allow it to make more correct and fewer incorrect decisions, and to do so at greater speed and lower cost.

In at least some respects, the law encourages such an interpretation of itself: We term it a system,[1] and we speak often of the reliability of evidence, the search for truth,[2] and the avoidance of error. We note the efficiency of the common law,[3] the unrestrained opportunities for taking judicial notice of legislative facts,[4] and the law's welcoming of new scientific and technical knowledge[5] of various kinds through expert witnesses. Moreover, many features of the law do lend themselves to empirical study, have been studied, and have yielded interesting and potentially useful new knowledge.[6] Nevertheless, numerous findings of empirical social science that may inform procedural choices have not been embraced by legal policy.

---



* Professor of Law, University of Iowa.

The author wishes to acknowledge the interdisciplinary research grant support provided by University House and the University of Iowa, which facilitated his work on this project.

1. For examples, consider the following titles: E. FARNSWORTH, AN INTRODUCTION TO THE LEGAL SYSTEM OF THE UNITED STATES (1983); J. STONE, THE LEGAL SYSTEM AND LAWYERS' REASONINGS (1964); JUSTICE SYSTEM JOURNAL.

2. "These rules shall be construed . . . to the end that the truth may be ascertained . . . ." FED. R. EVID. 102.

3. R. POSNER, ECONOMIC ANALYSIS OF LAW 12 (3d ed. 1986).

4. *See* FED. R. EVID. 201(a) advisory committee's note.

5. Giannelli, *The Admissibility of Novel Scientific Evidence:* Frye v. U.S. *A Half-Century Later*, 80 COLUM. L. REV. 1197, 1199 (1980).

6. To some extent, this two-part issue of *Law and Contemporary Problems* is an attempt to highlight such knowledge. Organizations such as the Federal Judicial Center, the State Justice Institute, and the National Center for State Courts exist in part to pursue and support the pursuit of such knowledge. For more information regarding the purposes and functions of these three organizations, see 28 U.S.C. §§ 620-639 (1982 & Supp. 1986); 42 U.S.C. §§ 10701-10713 (Supp. 1986); and *The National Center for State Courts*, STATE CT. J., Winter 1977, at 2, respectively.

If we put aside the possibility—even the likelihood—that some or many social science findings are premature, are based on insufficient data, or are just plain wrong (and by some magic the law "knows" that), we still find well understood phenomena being ignored or, worse, we find the law moving in a direction precisely opposite to the implications of the data. At other times the law embraces erroneous findings and non-expertise (my euphemism for well-intended quackery or outright charlatanism) and displays a determined need not to hear data that tend to reveal and correct the errors.

Of course, the principal purpose of the legal process is not to obtain correct answers; it is to resolve disputes. Finding facts is merely instrumental to the larger purpose of the legal process; it plays a supporting and not a leading role. The legal process is a truth-seeking enterprise only insofar as truth is needed to resolve disputes. And, inevitably, at some point, the processes of truth-finding could grow to exact costs (dollars, time, intrusiveness, and so on) that exceed any benefit the parties or society could gain from the effort. At some point the game is not worth the candle. The law must impose some limitation on the development of its truth-finding methods before the litigation system becomes bogged down in its own efforts to do its job well.

Moreover, the issues that usually present themselves include both cognitive conflicts that are best resolved through factfinding and distributional conflicts that require making other kinds of judgments.[7] This mixture may help explain some of the choices the law makes in structuring its procedures and deciding what kinds of evidence to admit. The consideration of facts depends upon their relevance, and their relevance is determined by the applicable substantive law. The resources the law is willing to devote to finding materially relevant facts may depend upon how pivotal those facts are to the case. In resolving disputes involving mixed questions of cognitive uncertainty and distributional conflict, still other considerations may dominate the process and the decisions reached. Therefore, it is essential to realize the important but limited part that factfinding plays in dispute resolution.

In addition, the dispute resolution process itself resides within a larger political system. Any dispute resolution process has to be satisfying to the disputants, including the losing parties, and to the larger society.[8] At various stages of our society's development, that may mean preferring one set of procedures rather than another, even if the favored rule reduces rather than increases the system's factfinding capability.

Given these observations, it is noteworthy that, if anything, the values of the larger culture and the legal culture seem over time to have moved toward an increased insistence on true as well as correct results, raising the importance of the factfinding component of the dispute resolution process.

---

7. Thibaut & Walker, *A Theory of Procedure*, 66 CALIF. L. REV. 541 (1978).

8. *See generally* J. THIBAUT & L. WALKER, PROCEDURAL JUSTICE: A PSYCHOLOGICAL ANALYSIS (1975).

We could resolve disputes expeditiously and unambiguously by tossing a coin to decide liability. Or we could submit our cases to an oracular examiner of chicken entrails. An answer would emerge. But such decision processes would quickly erode public confidence and would soon be abandoned. The evolution of a societal preference for reducing fact-based uncertainties has apparently been mirrored in the law by this century's increased use of discovery and more porous rules of evidence. And error, especially the demonstrable kind, has its own costs, both to immediate parties and to the perceived legitimacy of the legal system.

It is precisely in light of such considerations of the proper role to be played by factfinding in the legal process that some of the choices made by the law become most puzzling. With little imagination we could explain why the law might refuse to make use of a complex and costly factfinding tool that added only marginally to the factfinder's understanding, or even why the law would fail to avail itself of inexpensive and relevant knowledge. But it is hard to explain why—in a society that both places a high value on accuracy and efficiency and desires to keep political control out of the hands of technocratic experts—courts would make increasing use of sources of knowledge that add very little to—or, indeed, subtract from—the accuracy of legal factfinding. This practice becomes even more puzzling when the courts' measures increase the expense and duration of the legal process and place much discretion in the hands of experts rather than judges or juries. If the law is making such counterproductive choices—and I aim to show that it is doing so openly—then the legal process not only is failing to serve the public as well as it might, but it is also running the risk that once word gets around, public respect for the legal process will wane. In a sense, this article is an inquiry into whether these seemingly irrational choices can be explained by a deeper inspection of the functioning of our legal process. Perhaps there are broader and more important functions of the system that are served by choices that appear counterproductive when viewed in a narrow scope.

## II

## SOCIAL SCIENCE KNOWLEDGE AND THE ENHANCEMENT OF ACCURATE FACTFINDING

This section reviews a number of examples of social science findings that throw light on the quality of the law's factfinding and suggest improvements to the design of the law as a decisionmaking system in the sense of enhancing the accuracy of its findings.

### A. Evaluation of Overall Reliability and Validity

As a system of deciding disputes, the law aims to achieve reliability, at least, and validity, if possible. Thus, the law ought to strive to treat like cases

alike (reliability)[9] and, if possible, to make the correct decision on those similarly treated similar cases (validity).[10]

The problem of assessing validity presents a greater challenge in law than in any other field. Geologists can see if oil is where they think it is or if earthquakes strike where and when predicted; physicians can perform autopsies to test the accuracy of their diagnoses; clinical psychologists can look to actual behavior as a criterion against which to compare the accuracy of their tests. But the adjudicative factfinding part of the law operates only in a retrospective, historical mode, and can never truly test its findings against some external criterion.[11] Moreover, the law contains a central thread that is absent from comparable institutions that engage in similarly complex decisionmaking: the value of equal treatment. Patients may complain that they did not receive the best or most up-to-date care; they do not complain that they were not treated equally. Litigants can and do make such complaints against the legal system, and the system takes those complaints seriously.

The unusual difficulty of finding a criterion against which to test the correctness of trial outcomes and the special concern in the law for equal process lead to an emphasis on reliability, rather than validity, in evaluating the working of the law.[12]

How reliable are legal factfinders? One well-known source of data on this point is the work of Kalven and Zeisel.[13] In a number of jury trials, by comparing verdicts of civil juries with verdicts that judges said they would have rendered if the trials had been bench trials, it was found that judges and juries agreed on the verdict in 79 percent of the cases.[14] Even this study

---

9. In measurement theory, reliability is defined as the ability of a measure to produce the same result each time the same object or event is evaluated by the measuring device. J. MONAHAN & L. WALKER, SOCIAL SCIENCE IN LAW 43-45 (1985). The same notion is captured by the law's recognition of the need to treat like cases alike.

10. Validity is defined as the ability of a measure to measure what it purports to measure. J. MONAHAN & L. WALKER, *supra* note 9. To analogize to ideas familiar to lawyers, once we are satisfied that like cases are treated alike, we can ask whether the correct decisions have been reached on those similar cases.

11. Empirical science has a decisionmaking limitation that parallels the law's. The effects of one variable on another are never known with certainty, but only to the degree that assumptions of non-difference, or null hypotheses (the equivalent of presumptions in the law), can be rejected when sufficient data (evidence) are accumulated in an experiment to meet the required level of significance (standard of proof). In this way, the fundamental logic of hypothesis testing and legal factfinding are remarkably similar. The language is different, but the concepts and logic are strikingly parallel. The problem remains, however, that the two fields aim their respective factfinding machinery in different directions: the law toward bits of mundane and disputed history (adjudicative facts); social science toward repeatable phenomena (typically, legislative facts). Without turning to or reinventing social science methods or findings, the law cannot answer its questions about the efforts of alternative legal procedures.

12. *See* Walker, *Perfecting Federal Civil Rules: A Proposal for Restricted Field Experiments*, LAW & CONTEMP. PROBS., Summer 1988, at 67.

13. Unfortunately, these two researchers never fully analyzed their laboriously collected civil data, and now those data have been lost. Only the criminal data have been analyzed and reported in detail, in H. KALVEN & H. ZEISEL, THE AMERICAN JURY (1966). *See infra* note 14.

14. Judges and juries would both find for plaintiffs in 44% of cases, both for defendants in 35% of cases, judge for plaintiffs and jury for defendants in 10% of cases, and judge for defendants and jury for plaintiffs in 11% of cases. Kalven, *Dignity of the Civil Jury*, 50 VA. L. REV. 1055, 1065 (1964).

provides only a proxy answer to the question of reliability; these data do not tell what happens if the same case is "measured" repeatedly by the legal system by submitting the same case repeatedly to the system.[15] But let us take what we have to be the estimate of the system's reliability.

Is 79 percent agreement good or bad? It is less than perfect, to be sure. And the lower end of the scale of possibilities is not 0 percent but 50 percent—the level of agreement that would be reached if judge and jury tossed a coin to decide their respective verdicts.[16] To be below 50 percent agreement would be astonishing as well as disturbing. Perhaps the best standard of evaluation is to compare the courts with makers of other complex decisions. Table 1 reproduces data provided by Diamond.[17] The civil verdict agreement rate of 79 percent is at least equal, and sometimes superior, to the agreement rate of National Science Foundation ("NSF") and National Academy of Sciences ("NAS") peer reviewers, employment interviewers, psychiatrists and physicians making diagnostic decisions, and judges making sentencing decisions. Thus, the legal process produces verdicts whose reliability sets a standard to which physicians, scientists, and business people might aspire.

Despite this relatively impressive performance, the general public and even lawyers commonly believe that one never knows what a judge or jury will do. Unreliability of decisionmaking is often cited as one of the major problems with the civil justice system[18] that necessitates reform. Based on these data, one might think that the public would believe that one never knows what physicians/NSF or NAS scientists/employment interviewers and so on will do, and that the public would be pressing for reforms of those institutions.

---

15. And this cannot really be the end of an inquiry about the reliability of the legal machine. The level of judge-jury agreement found by Kalven and Zeisel could have been achieved because the same legally irrelevant, or even legally impermissible, ingredient determined the outcome (for example, the occupation of the plaintiff). And reliance on such irrelevant data would violate, not effectuate, the doctrine of equal protection. Thus, the statistical evidence of reliability might be grounded on practices that are anti-reliable as between parties who are *legally* similarly situated.

16. The combinations of coins would be: HH or TT 50% of the time, and HT or TH 50% of the time. That exhausts the possibilities.

17. Diamond, *Order in the Court: Consistency in Criminal Court Decisions*, in 2 THE MASTER LECTURE SERIES, PSYCHOLOGY AND THE LAW 119, 125 (C. Scheirer & B. Hammonds eds. 1983).

18. One of the central themes of the literature of the "civil liability crisis" in the past few years has been unpredictable and inconsistent decisionmaking by the civil justice system. For example, the Model Health Care Provider Liability Reform Act section 2(a)(3) states that one of the principal problems it confronts is "the unpredictability of medical malpractice awards," and section 2(b) of the Act states that one of the benefits it promises is "to establish greater predictability." The *Description*, which accompanies the Act, states, "Specifically, the current tort system, frequently described as a litigation lottery, lacks predictability." U.S. Dep't Health & Hum. Servs., Model Health Care Provider Liability Reform Act §§ 2(a)(3), (b) (1988).

Some of the empirical research in this area focuses on dichotomous liability verdicts; other research focuses on continuous awards. More research exists on the former than the latter. Kalven, *supra* note 14, reports that while civil juries and judges had a high level of agreement on liability, the juries reached award amounts that averaged 20% higher than those judges would have awarded. More discussion of awards appears in the next section.

TABLE 1

RELIABILITY OF COMPLEX DECISIONMAKING IN SEVERAL FIELDS

INTERJUDGE CONSISTENCY IN COMPLEX HUMAN JUDGMENTS

| Decisionmakers | Stimulus | Decision | Rate of agreement between 2 judges (%) |
|---|---|---|---|
| NSF versus NAS peer reviewers | 150 grant proposals submitted to NSF | To fund or not to fund (half funded by NSF) | 75 |
| 7 employment interviewers | 10 job applicants | Ranked in top 5 or in bottom 5 | 70 |
| 4 experienced psychiatrists | 153 patients interviewed twice, once by each of 2 psychiatrists | Psychosis, neurosis, character disorder | 70 |
| 21-23 practicing physicians | 3 patient-actors with presenting symptoms (Doctors could request further information and could order and receive test results). | Diagnosis: correct or incorrect<br>Probability of agreement (both correct or both incorrect)[a] | 67,77,70<br>55,65,57 |
| 3,576 judge-jury pairs | 3,576 jury trials | Guilty or not guilty | 78 |
| 12 federal judges | 460 presentence reports (at sentencing council) | Custody or no custody | 80 |
| 8 federal judges | 439 presentence reports (at sentencing council) | Custody or no custody | 79 |

[a] Inflated because physicians could also be inaccurate in different ways.

Source: Diamond, *Order in the Court: Consistency in Criminal Decisions*, in 2 THE MASTER LECTURE SERIES, PSYCHOLOGY AND THE LAW 119, 125 (C. Scheirer & B. Hammonds eds. 1983).


One might think, further, that just as all of these other institutions wish to improve the reliability of their decisions,[19] so would the law seek to adopt reforms, or at least some fine-tuning, that would enhance the reliability of the civil justice system's decisionmaking.[20]

## B. Juries

Research on judge-jury reliability represents only the beginning of the study of juries as a component of the legal process. This is the area of the most extensive social-psychological research germane to the legal process. It

19. Reliability is a major concern in psychology, medicine, and other fields. *See, e.g.*, Bates & Saxe, *Reimbursement for Psychotherapy: Linking Efficacy Research and Public Policymaking*, 38 AM. PSYCHOLOGIST 918 (1983); U.S. CONGRESS, OFFICE OF TECHNOLOGY ASSESSMENT, THE QUALITY OF MEDICAL CARE: INFORMATION FOR CONSUMERS 3-7 (1988).

20. I will discuss particular examples of the law moving away from reliability in areas such as jury size and polygraph use later in this article. Another instance of the law ignoring empirical research is illustrated by the widespread belief in a "litigation explosion" among legislators despite evidence to the contrary. *See generally* Galanter, *The Day After the Litigation Explosion*, 46 MD. L. REV. 3 (1986); Saks, *If There Be a Crisis How Shall We Know It?*, 46 MD. L. REV. 63 (1986).

includes studies of jury composition effects,[21] competence,[22] complex civil litigation,[23] effects of unanimous versus quorum verdicts,[24] persuasion,[25] communication and influence within the jury,[26] the decisionmaking process,[27] and a diverse collection of studies of how different kinds of information affect (or do not affect) juries' decisions.[28]

I will focus here on a simple and familiar issue that is rather well researched and understood, and on which the Supreme Court has spoken (repeatedly)[29] and legislatures have acted: the size of the civil jury. Recall our discussion at the beginning of this article on one of the most fundamental concerns of the legal decisionmaking process: its reliability, in the sense of producing consistent and predictable results. What is the relationship between the size of the jury and the reliability of the decisions it makes?

This is one of the easier questions for a statistician or social scientist to answer. Moreover, the answer has been recognized in an opinion of the United States Supreme Court,[30] even as that opinion again affirmed the right of states to reduce juries to groups as small as six. The general relationship has been summarized by Hans and Vidmar:

> A statistical analysis reveals that over two-thirds of the twelve-person juries will have damage awards close to the community average, compared to just half of the six-person juries. The six-person juries are four times as likely to have extremely low or extremely high average damage awards. Hence, the twelve-person jury should provide a more accurate and a more reliable reflection of the community's assessment.[31]

In fact, we can calculate the resulting increase in error[32] with precision. All things being equal, a reduction of group size by one-half increases variability (error, unreliability) by a factor of 1.41.[33]

---

21. *See generally* Saks, *The Limits of Scientific Jury Selection: Ethical and Empirical*, 17 JURIMETRICS J. 3 (1976).

22. *See generally* H. KALVEN & H. ZEISEL, *supra* note 13; R. HASTIE, S. PENROD & N. PENNINGTON, INSIDE THE JURY (1983).

23. *See generally* Lempert, *Civil Juries and Complex Cases: Let's Not Rush to Judgment*, 80 MICH. L. REV. 68 (1981); M. SAKS, SMALL-GROUP DECISION MAKING AND COMPLEX INFORMATION TASKS (1981).

24. *See generally* M. SAKS, JURY VERDICTS (1977); V. HANS & N. VIDMAR, JUDGING THE JURY 171-75 (1986).

25. *See* M. SAKS & R. HASTIE, SOCIAL PSYCHOLOGY IN COURT 101-10 (1978).

26. *Id.* at 72-99.

27. *See generally* R. HASTIE, S. PENROD & N. PENNINGTON, *supra* note 22.

28. *See* review in V. HANS & N. VIDMAR, *supra* note 24, at 113-63, 179-217.

29. *See generally* Ballew v. Georgia, 435 U.S. 223 (1978); Colgrove v. Battin, 413 U.S. 149 (1973); Williams v. Florida, 399 U.S. 68 (1970).

30. Ballew v. Georgia, 435 U.S. at 232-33.

31. V. HANS & N. VIDMAR, *supra* note 24, at 167.

32. At least that portion that is attributable to the statistical effects.

33. The standard error associated with samples of any size is given by the formula:

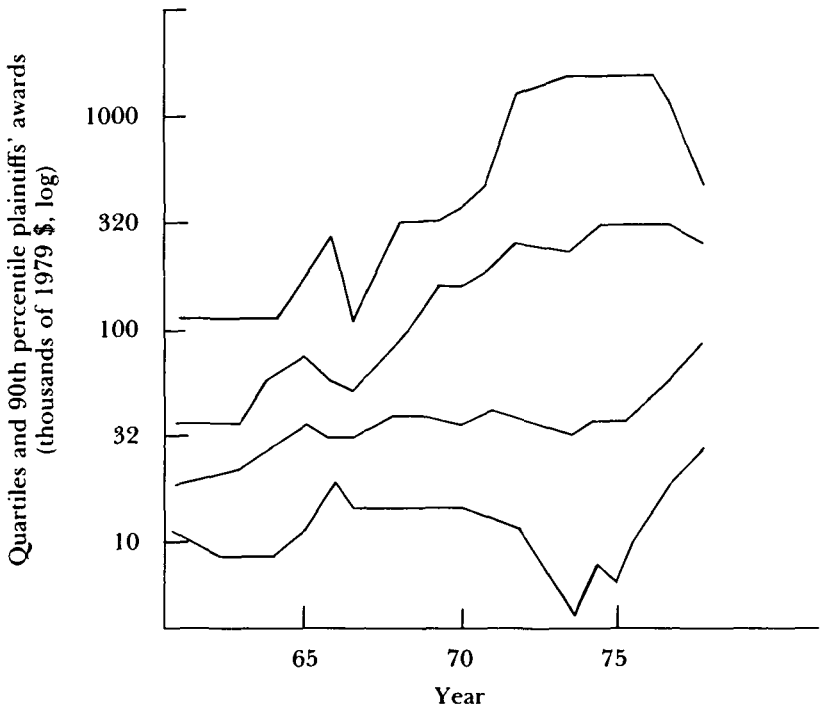
$$\sigma_e = \frac{\sigma}{\sqrt{N}}$$

If the $N$ is reduced, then $\sigma_e$ will increase; if $N$ is reduced by one half, $\sigma_e$ will increase by the square root of two, that is, 1.41. This applies to dollar awards. A similar result occurs for a dichotomous variable such as verdicts.

Some confirmation of this effect comes from recent studies of civil jury awards. Figure 1 represents jury awards in medical malpractice cases in Cook County, Illinois.[34] The divergence between the 10th and 90th percentiles in awards plotted over time depicts the increase in variability. While this may be due to something other than reduction in jury size, it certainly is consistent with expectations.[35]

FIGURE 1

RAND DATA ON CHICAGO MALPRACTICE TRIAL AWARDS




Source: M. PETERSON & G. PRIEST, THE CIVIL JURY: TRENDS IN TRIALS AND VERDICTS, COOK COUNTY, ILLINOIS, 1960-1979 (R-2881-ICJ), at 35 (1982) (Institute for Civil Justice, Rand Corp.).

Knowledge of the relationship between the size of the decisionmaking group and reliability provides the legal system with a means to increase or decrease the variability in verdicts and awards with some precision. To reduce the range of random variation in awards, we need only increase the

34. M. PETERSON & G. PRIEST, THE CIVIL JURY: TRENDS IN TRIALS AND VERDICTS, COOK COUNTY, ILLINOIS, 1960-1979 (1982) (published by the Rand Corp.). We do not know the actual size distribution of juries in Chicago; such studies are still needed. Illinois law provides that where claims for damages do not exceed $15,000, the trial shall be a jury of six unless one of the parties demands trial by a jury of 12. Cases where the damages claim exceeds $15,000 are to be tried before juries of 12 unless the parties agree to a lesser size. ILL. ANN. STAT. ch. 110, para. 2-1105 (Smith-Hurd Supp. 1988).

35. Research is required to test competing explanations for the divergence. Almost certainly, a substantial portion of the increase in variability is due to the reduction in jury size.

size of the decisionmaking body;[36] to increase the error, we need only reduce the size of the group.

## C. Forensic Science Proficiency Testing

Although not usually thought of as social science research, I would include in the present discussion the proficiency testing studies recently undertaken in the forensic sciences.[37] I think these studies may be properly classified as social science research because the behavior of forensic scientists is the subject of inquiry. The proficiency of forensic scientists is tested by sending them known quantities of hair, blood, fingerprints, handwriting, and so on to examine and analyze. In an actual case, one can never know with certainty whether the known sample and the questioned sample came from the same source. However, in these tests, the source of each sample is known by the manufacturer of the test materials. Thus, the performance of forensic scientists can be tested by comparing their answers against the known characteristics and origins of the materials tested. The result provides both reliability data (the extent to which forensic scientists agree with each other in answering the same problem) and validity data (the accuracy of their answers).[38]

The forerunner of the present program of proficiency testing was conducted by Peterson, Fabricant, Field, and Thornton.[39] Their findings, summarized in Table 2,[40] provide data on a wide array of forensic science subdisciplines. I have developed a special curiosity about the reliability and validity of document examiners' conclusions concerning whether a known and a questioned writing were written by the same person.[41]

---

36. This theory has its limits, of course, because we are dealing here not only with a statistical phenomenon, but a social one as well. Above a certain size, the social processes of group communication and decisionmaking would overwhelm the several advantages of group size, thereby degrading the group's performance. *See generally* M. SAKS, *supra* note 23; I. STEINER, GROUP PROCESS AND PRODUCTIVITY 83 (1972).

37. For a detailed analysis of these testing programs, see generally Lucas, Leete & Field, *An American Proficiency Testing Program*, 27 FORENSIC SCI. INT'L 71 (1985).

38. Proficiency studies can be defined to set more demanding levels of acceptable performance: Not only must one get the "right" answer, one must do so by use of the proper tests. For this reason, the figure of 71.2% unacceptable responses in Table 2 may be interpreted to overstate the invalidity of the findings of blood testing.

39. J. PETERSON, E. FABRICANT, K. FIELD & J. THORNTON, CRIME LABORATORY PROFICIENCY TESTING RESEARCH PROGRAM (1978) (U.S. Gov't Printing Office).

40. *Id.* at 251.

41. *See generally* Risinger, Denbeaux & Saks, *Exorcism of Ignorance as a Proxy for Rational Knowledge: The Lessons of Handwriting Identification "Expertise,"* 137 U. PA. L. REV. 731 (1989).

## TABLE 2

## CRIME LABORATORY PROFICIENCY TESTING PROGRAM

PERCENTAGES OF LABORATORIES REPORTING RESULTS OF "UNACCEPTABLE PROFICIENCY"

$$\frac{\text{Number "unacceptable" responses}}{\text{Number of laboratories responding with data}} \times 100 = \text{Percent "Unacceptable"}$$

| Sample Number | Sample Type | Number of Labs Responding With Data | Number of "Unacceptable" Responses | % of Laboratories Submitting "Unacceptable" Responses |
|---|---|---|---|---|
| 1 | Drugs | 205 | 16 | 7.8 |
| 2 | Firearms | 124 | 35 | 28.2 |
| 3 | Blood | 158 | 6 | 3.8 |
| 4 | Glass | 129 | 6 | 4.8 |
| 5 | Paint | 121 | 24 | 20.5 |
| 6 | Drugs | 181 | 3 | 1.7 |
| 7 | Firearms | 132 | 7 | 5.3 |
| 8 | Blood | 132 | 94 | 71.2 |
| 9 | Glass | 112 | 35 | 31.3 |
| 10 | Paint | 111 | 57 | 51.4 |
| 11 | Soil | 93 | 33 | 35.5 |
| 12 | Fibers | 120 | 2 | 1.7 |
| 13 | Physiological Fluids (A & B) | 129 | (A) 3<br>(B) 2 | (A) 2.3<br>(B) 1.6 |
| 14 | Arson | 118 | 34 | 28.8 |
| 15 | Drugs | 143 | 26 | 18.2 |
| 16 | Paint | 103 | 35 | 34.0 |
| 17 | Metal | 68 | 15 | 22.1 |
| 18 | Hair (A,B,C,D,&E) | 90 | 45<br>25 | Dog (A)50.0<br>Cat (B)27.8<br>Deer (C)54.4<br>Cow (D)67.8<br>Mink (E)35.6 |
| 19 | Wood | 65 | 14 | 21.5 |
| 20 | Questioned Documents (A&B) | 74 | 4<br>14 | (A) 5.4<br>(B)18.9 |
| 21 | Firearms | 88 | 12 | 13.6 |

Source: J. PETERSON, E. FABRICANT, K. FIELD & J. THORNTON, CRIME LABORATORY PROFICIENCY TESTING RESEARCH PROGRAM 251 (1978).

The data on document examiner proficiency have been summarized as follows:

> A rather generous reading of the data would be that in 45% of the reports forensic document examiners reached the correct finding, in 36% they erred partially or completely, and in 19% they were unable to draw a conclusion. If we assume that inconclusive examinations do not wind up as testimony in court, and omit the inconclusive reports, and remain as generous as possible within the bounds of reason, then the most we can conclude is this: Document examiners were correct 57% of the time and incorrect 43% of the time.
>
> But let us turn to more meaningful readings of the aggregate data . . . . Disguised handwriting fooled them all and forged printing fooled two-thirds of those who hazarded an opinion about it . . . .
>
> If a correct answer consists of a report containing correct conclusions returned pursuant to requested and submitted test materials, then of the total submissions to

> laboratories in the 1984 through 1987 tests, only 18% gave wholly accurate responses . . . .
>
> Finally, consider the possible effect on any aggregate conclusions of the fact that . . . only a fraction [of police laboratories] even ordered test materials in the first place. It is at least arguable that, by self-selection, the sample is inherently biased in favor of the more conscientioius and capable practitioners to begin with. If this is true, the reported results would overstate the accuracy of the handwriting examination field generally.[42]

These data show that document examiners as a group may not have quite the expertise that they would like courts to believe they possess, that the outcome of the case may depend upon which examiner has been called to testify, and that the level of performance varies with the precise nature of the task at hand. These data may allow a jury to make a more informed estimate of the weight to be accorded these witnesses, but they do not yet help a court to know whether the experts add anything to what jurors might be able to accomplish by their own direct comparison of the writings.[43]

While the data of proficiency studies are somewhat controversial with respect to how well they inform us about the reliability and validity of the various forensic science sub-disciplines represented,[44] they certainly are relevant to the question of the probity of forensic science testimony. Whether the proficiency data somewhat overstate or understate the actual level of validity, they strongly suggest that the actual level is almost certainly lower than the average factfinder's assumption of that level,[45] and the basic insight they give a factfinder about the possibility of error is itself sobering. The

---

42. *Id.* at 747-49.

43. Without comparative data, we cannot know if laypersons would do worse, as well as, or better than document examiners.

44. At the basis of the controversy is an uncertainty as to exactly who is taking the test and how the interpreter is treating it. Does a laboratory assign the test to a beginner or to the most experienced person? Does the examiner give it less or more attention than a sample from an active crime investigation would receive? The answers to these questions are unknown—and the forensic science discipline may have an interest in obtaining the proficiency data while keeping the interpretation of it ambiguous.

As discussed in Risinger, Denbeaux & Saks, *supra* note 41, the earliest attempts to study the proficiency of document examiners found them to be so elusive and resistant that it simply was not possible to subject their claimed abilities to controlled study. Modern proficiency studies, which include most or all laboratory forensic sciences, are carefully shrouded in anonymity. The test producers invite laboratories to participate; those who choose to participate do not send in identifying information with their answers; to find out how well they did, each laboratory can compare the answers they gave to an answer key and report of how labs in the aggregate did. The test producers know how well those laboratories that participated did, in general, but they cannot know who sent back answers, who did not, and who did how well. Even these aggregate results are not published or otherwise made accessible to the public, courts, lawyers, or even to forensic scientists who did not participate.

It seems clear that forensic scientists have been unusually slow to subject themselves to such testing, and now that they have begun to do it, they have gone to considerable trouble to keep the results quiet and anonymous (so that no given practitioner or laboratory has to be answerable for its performance). One can only speculate on the advantages this creates. Here is one speculation: If one is working in a discipline whose evidentiary offerings go largely unchallenged by opposing lawyers and are blindly accepted by judges and juries, then one has nothing to gain and everything to lose by subjecting one's claimed abilities to tests and making the results known.

45. Saks & Wissler, *Legal and Psychological Bases of Expert Testimony*, 2 BEHAVIORAL SCI. & L. 435, 443-45 (1984).

interesting question, then, becomes whether these data are proffered to courts and, if so, whether they are admitted or acknowledged through judicial notice.[46]

Although courts generally are open to forensic science evidence, they do not appear eager to admit evidence about the *reliability and validity* of forensic science evidence—that is, evidence about the evidence. Some regard it as simply not relevant (proficiency data are about other forensic scientists in other places at other times doing other cases, not about this forensic scientist here today in this case); others find more elaborate explanations for its inadmissibility.[47]

Walker and Monahan, however, suggest that the use of data about general proficiency to make an individualized judgment has become more common, and they find no barrier to its admissibility in the usual doctrines of evidence law. They term this kind of use a "framework" use of social science data. That is, the information about a forensic science's general proficiency provides a framework for assessing the probity or the weight to be given the testimony of the forensic scientist on the stand in a given proceeding.[48] Presumably, taking into account the limited reliability and validity of the proffered evidence would enhance the reliability and validity of the judicial factfinding process, and vice versa.

## D. Probability Data and Behavioral Decision Theory

The use of background data to go from the general (other forensic scientists) to the particular (this expert witness) requires some probabilistic inference by the factfinder. Whenever judges or jurors have to make intuitive probability guestimates, they fill in missing information with implicit assumptions. And of course they arrive at their estimates without the benefit of Bayes' theorem (or any other computational aid), but only with their intuitions about what to make of the information. The law has important

---

46. The apparent answer is that even as that research sent shock waves through the forensic science community, it evoked barely a yawn from the legal community. A Westlaw search for appellate cases citing the proficiency testing literature, in particular the study by J. PETERSON, E. FABRICANT, K. FIELD & J. THORNTON, *supra* note 39, turned up exactly two cases: State v. Flynn, 202 N.J. Super. 215, 220 n.2, 494 A.2d 350, 353 n.2 (1985), and Chappee v. Commonwealth, 659 F. Supp. 1220, 1221 n.2 (D. Mass. 1987). Both of these courts appear to have taken the findings seriously and cited them as support for rulings requiring greater scrutiny of forensic science experts.

47. For example, some courts adopt a policy that only one who is qualified as an expert in the particular forensic science field is qualified to present the proficiency data. By this analysis, ignorance about the proficiency data in one's field strengthens the barrier to its admission and strengthens the basic testimony.

The author's source for this point is his own experience in court in ongoing litigation that cannot be identified, so unfortunately the reader will have to take his word for it that he has seen a trial judge rule as described.

48. Walker & Monahan, *Social Frameworks: A New Use of Social Science in Law*, 73 VA. L. REV. 559 (1987). By their analysis, though, the general proficiency data ought to be dealt with as a matter of law, so that the court may take judicial notice of it, receive briefs and arguments on it, and instruct the jury about it. The court would thus have a duty to notice such data and to advise the jury as to it. Interestingly, the courts still seem to prefer to insulate the jury from knowledge of the limitations of forensic scientists and other witnesses. *See supra* note 42.

choices to make in deciding what to give jurors and what not to give them. On the one hand, the law wants them to have enough information to reach a rational judgment. On the other hand, it does not want them to receive information that will mislead or confuse them. Determining which information is helpful and which is harmful is a task in which judges can receive some help from the growing body of empirical research that is carried out to discover how humans process probabilistic information and draw intuitive inferences from that information.

Here it is useful to focus on a few examples showing how that knowledge can assist courts in deciding whether to admit certain kinds of evidence. First, courts have to decide if probability evidence is to be admitted at all. If it is not deemed irrelevant solely because of its probabilistic, aggregate character,[49] then the court may have to decide if it is sufficient to support a verdict for a plaintiff. This is, of course, an area of considerable confusion for the law. Sometimes such evidence is disallowed as not relevant,[50] and sometimes it is disallowed as being unable on its own to meet the plaintiff's burden of persuasion.[51] Other times the courts gladly welcome what is in form the same kind of information.[52] They almost always will allow the evidence in if it is filtered through a case-specific expert witness who is employing the evidence to reach the opinions being testified to.[53] Probability data can serve at least two basic purposes in legal factfinding. Such data can be informative as primary evidence (for example, the probability of lung cancer after exposure to asbestos is increased five times; after exposure to asbestos plus tobacco,

---

49. *See, e.g.*, Tribe, *Trial by Mathematics: Precision and Ritual in the Legal Process*, 84 HARV. L. REV. 1329, 1344-50 (1971); Saks & Kidd, *Human Information Processing and Adjudication: Trial by Heuristics*, 15 LAW & SOC'Y REV. 123, 124-25 (1981); MCCORMICK ON EVIDENCE § 210 (E. Cleary 3d ed. 1984); Walker & Monahan, *supra* note 48, at 572-75.

50. MCCORMICK ON EVIDENCE, *supra* note 49, §§ 209, 210.

51. Smith v. Rapid Transit, 317 Mass. 469, 470, 58 N.E.2d 754, 755 (1945): "[I]t is not enough that mathematically the chances somewhat favor a proposition to be proved; for example . . . the fact that only a minority of men die of cancer [would not] warrant a finding that a particular man did not die of cancer." *But see* H. HART & J. MCNAUGHTON, EVIDENCE AND INFERENCE IN THE LAW 54-55 (1958):

> Consider the formula that in a civil case the facts must be determined on a more-likely-than-not basis. In the first place, the probabilities are determined in a most subjective and unscientific way: the trier of fact simply asks itself which of the contesting contradictory propositions according to the trier's limited experience more nearly squares with the evidence. In the second place, the law refuses to honor its own formula when the evidence is coldly "statistical."

52. These instances include unexceptional cases involving forensic science and medical testimony as well as discrimination and toxic tort cases. *See generally* E. IMWINKELRIED, SCIENTIFIC & EXPERT EVIDENCE (2d ed. 1981) (forensic and medical evidence); D. BALDUS & J. COLE, STATISTICAL PROOF OF DISCRIMINATION (1980); E. GREER & W. FREEDMAN, TOXIC TORT LITIGATION (1989).

53. The apparent irony here is that this procedure allows the witness to become the decisionmaker on that point. The case-specific witness could simply report to the factfinder the relevant facts and then report the underlying aggregate scientific findings. Having made that explicit, the expert could leave the stand and allow the factfinder to draw the appropriate conclusion. The expert could even suggest (by opinion) what conclusion one is to draw from the adjudicative facts combined with the aggregate data framework. However, the court wants just the expert's final decision. The underlying data can come out if one attorney or the other wants it to. Does the law actually have a different goal, namely, the exorcism of ignorance? *See generally* Risinger, Denbeaux & Saks, *supra* note 41.

sixty times), or it can be useful to provide the baserates that give meaning to primary evidence—a matter to be discussed shortly.

The law both embraces and rejects probability evidence based upon aggregate data. Such information is what modern scientific knowledge is made of. If courts embrace such knowledge, it is said that "when statistics speak, courts listen," or that without such data there would be no way to resolve the issues before the courts.[54] And, of course, these courts are right. When courts reject probability evidence, they point out that whatever the aggregate data show, those data provide no assurance that what is true in general is true in the particular case before the court; one cannot decide the particular with certainty on the basis of inferences drawn from general probabilities.[55] And, of course, these courts are right as well.

The sorting out of these contradictions is in itself a major challenge for judges and legal scholars. The examples I wish to discuss here involve more practical concerns about presenting the decisionmaker with the right data in the best form so as to lead to an intelligent appraisal of the facts of a situation.

In making sense of most adjudicative assertions, baserate information is helpful, perhaps necessary, and by definition is statistical and probabilistic. One example from a criminal context will make this clear. Suppose a forensic scientist takes the stand to report that microscopic glass fragments were found in the defendant's outer clothing, and these fragments are indistinguishable in chemical composition and refraction from the glass window that was broken to gain entry into a victim's home. A factfinder might well conclude that this is highly identifying information. A court might well hesitate to allow baserate information about the population of glass fragments, as tending to cloud the picture. But a factfinder can find the initial expert testimony helpful only if it is also assumed that the likelihood of finding a "match"[56] is low unless the defendant was at the victim's home.

In the absence of testimony on these points, how do the factfinders fill in the missing information? The answer can only be that they make it up; they guess. The testimony about the glass fragments is diagnostic, or identifying, only to the degree that the baserate of glass fragments is low. In a study of this question, criminalists collected outer clothing from a dry cleaner and examined it for glass fragments. They found that 60 percent of the outer garments contained fragments that were the same in chemical composition and refraction as the most common type of window glass. This baserate alone means that any defendant selected at random stands a 60 percent chance of being found to possess such fragments. Thus, in this instance, the baserate shows the weak diagnosticity of the evidence. In other instances, it might

54. *See* D. BARNES, STATISTICS AS PROOF 1 (1983); MCCORMICK ON EVIDENCE, *supra* note 49, §§ 210-11.

55. *See* MCCORMICK ON EVIDENCE, *supra* note 49, § 210.

56. I am using this term for the purposes of convenience and overstatement. Conscientious forensic scientists avoid the term "match," and instead describe the state of affairs more correctly as "consistent with the evidence" or, when the probabilities are weak, as "not inconsistent with the evidence."

show that the evidence is far more diagnostic than the factfinders might assume it to be.

Most evidence, and certainly scientific evidence, rests on just such probabilistic inferences. Sometimes courts reject such evidence as a basis of liability.[57] But in essence these courts are rejecting only the explicit data that would allow the factfinder to draw an informed inference. To exclude the baserate data does not make the inference non-probabilistic; it only prevents the inference from being informed by the data and requires the factfinders to substitute their guess for the researcher's data.[58]

Now, assuming a court will allow such baserate evidence, the next problem concerns the form it should take. Kahneman and Tversky[59] have shown that the way a problem is "framed" for a decisionmaker affects decisions based on identical information.[60]

Depending on the form in which the data are presented, the most basic statement of the same information can push the decisionmaker to err in one direction or another. Thompson and Schumann[61] have demonstrated the following: Suppose an expert witness supplies information on the degree to which a defendant and a perpetrator "match" on some characteristic. When such data are presented in the form of conditional probabilities, an *overestimate* is made about the diagnosticity of the information (what Thompson and Schumann call the "Prosecutor's Fallacy").[62] A conditional probability would take the form: "There is only a 2 percent chance the defendant's hair would be indistinguishable from that of the perpetrator if he were innocent." If the same information is presented as percentages of a population, the factfinder *underestimates* the probativeness of the evidence (the "Defense Attorney's Fallacy").[63] This would take the form: "Only 2 percent of people have hair that would be indistinguishable from that of the defendant, and therefore in a

---

57. *See* MCCORMICK ON EVIDENCE, *supra* note 49, § 210.

58. *See* Saks & Kidd, *supra* note 49, at 145-48.

59. Kahneman & Tversky, *Choices, Values and Frames*, 39 AM. PSYCHOLOGIST 341 (1984).

60. Here is an example of framing from Kahneman and Tversky:

> Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows:
>
> If Program A is adopted, 200 people will be saved.
>
> If Program B is adopted, there is a one-third probability that 600 people will be saved and two-thirds probability that no people will be saved.
>
> Which of the programs would you favor?
>
> Even though the expected saving of life in the two programs is identical, more than two-thirds of people presented with this problem favor Program A. By framing the problem in reverse, so that the number who die is stated (in Program A 400 will die; in Program B there is a one-third chance that no one will die and a two-thirds probability that 600 people will die), preferences for the programs reverse and Program B is preferred.

*Id.*

61. Thompson & Schumann, *Interpretation of Statistical Evidence in Criminal Trials: The Prosecutor's Fallacy and the Defense Attorney's Fallacy*, 11 LAW & HUM. BEHAV. 167 (1987).

62. *Id.* at 171.

63. *Id.*

city of 1,000,000 people, there would be approximately 20,000 such individuals."

Related studies of hindsight and attribution phenomena raise concerns about whether factfinders can adopt the designated burdens of proof and weigh evidence in the manner the law prescribes or expects. Hindsight findings show that after people have learned the outcome of an incident, that knowledge contaminates their assessment of the probability that the injury would result from the actions that preceded it. They consider the actual outcome to have been the expected outcome, even when its likelihood may have been remote, and they are unaware of the effect of the outcome information on their thinking.[64] Thus, the test for tort negligence, which directs factfinders to consider the behavior of the defendant ex ante, may be a test that human factfinders have great difficulty performing well—with the effect being implicitly to shift the burden of persuasion onto tort defendants.

Similarly, studies of the attribution of responsibility show that the more serious the injurious outcome, the more motivated observers are to attribute responsibility to someone, be that the plaintiff or the defendant.[65] The point is that random happenstance becomes a decreasingly acceptable attribution as the injury becomes more severe. The studies that demonstrate this effect present identical actions by the defendant, sometimes resulting in serious and sometimes in minor damage. The magnitude of the injury is relevant only to damages; the attribution of liability should be informed only by the defendant's conduct. Yet, for human decisionmakers, the two are more connected psychologically.[66] The greater the harm done, the greater the tendency of factfinders to attribute fault to the conduct of someone (usually the defendant).

Avoidance of such problems might require some restructuring of the way evidence is presented at trial. For example, the system can avoid these hindsight and attribution problems to the extent that it can bifurcate trials, withholding information about the magnitude of injuries and saving that information entirely for a separate damages phase of a trial. To do this completely would be difficult, if not impossible, however, since the mere fact that the trial is taking place tells factfinders that something did not turn out well, and that some non-trivial injury occurred.

## E. Polygraph

The polygraph has a relatively long past but only a short history. The notion that involuntary physiological changes are correlated with the awareness that one is lying is an idea that goes back centuries—at least to the

---

64. *See, e.g.*, Fischoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXPERIMENTAL PSYCHOLOGY: HUM. PERCEPTION & PERFORMANCE 288, 297-98 (1975).

65. *See, e.g.*, Walster, *Assignment of Responsibility for an Accident*, 3 J. PERSONALITY & SOC. PSYCHOLOGY 73 (1966).

66. It seems that the same connection holds for judges in deciding whether burdens of proof have been met. *See* H. HART & J. MCNAUGHTON, *supra* note 51, at 54-55. Judges, too, are human.

ancient Chinese and Bedouin Arabs.[67] While the polygraph itself has existed for most of this century, serious empirical research on the reliability and validity of the polygraph examination procedure, of which the hardware and electronics are only a part, began less than two decades ago.[68] A recent review of that literature by the Office of Technology Assessment ("OTA") provides the data in Tables 3A and 3B. When used in specific crime investigation, the polygraph performs at the level of validity shown in the tables.[69] The test's performance varies from study to study, but together the studies provide an approximate sense of the technique's ability to classify people as uttering what they believe to be truths or lies in specific crime situations. Used proficiently,[70] the polygraph technique is a clever tool which, like so many human inventions, has the capability to improve upon our unaided skills.

When we compare the ability of laypersons to detect lying, which is discussed in the next section, with the ability of well-conducted polygraph examinations, there appears to be no contest. Despite all of their experience and intuition, people are not skilled in the unaided detection of lying. With all of its limitations,[71] the polygraph examination process, in capable hands, errs less often.[72] Yet, the law's attitude toward this means of credibility assessment is clear. The right and power of jurors and judges to assess witness credibility intuitively is strongly protected and preserved—despite the fact that demeanor adds little to transcripts in terms of accuracy, and despite the fact that some of the very cues that factfinders are expected to rely upon actually *reduce* the accuracy of their assessments. Nevertheless, the law resists

---

67. J. BRIGHAM, SOCIAL PSYCHOLOGY 458 (1986).

68. This line of research was pioneered by David Raskin. He and David Lykken, known for his criticisms of the polygraph method and for his alternative questioning method (the Guilty Knowledge Test), have been at odds through most of the history of empirical research on polygraph testing. *See* L. SAXE, D. DOUGHERTY & T. CROSS, SCIENTIFIC VALIDITY OF POLYGRAPH TESTING: A RESEARCH REVIEW & EVALUATION—A TECHNICAL MEMORANDUM 29-43 (1983) (U.S. Congress, Office of Technology Assessment, OTA-TM-H-15).

69. Although these data appear rather impressive, it is only fair to note that the report from which they come is filled with caution about the risks of error. *Id.* at 4. As we shall see, if the standard is comparison with alternative sources of assessment that the law otherwise relies on, there seems little question that polygraph examination, employed proficiently (no small worry in itself, as we should appreciate from the above discussion of forensic science proficiency), is the easy winner, even with its imperfections.

70. Again, the caution word: proficient.

71. The most important limitation is that the data support its use only for the investigation of specific events which have occurred, not for employment screening, periodic security checks, and similar uses. The second limitation is that examiners must be proficient and blind to other information about the case under investigation. L. SAXE, D. DOUGHERTY & T. CROSS, *supra* note 68, at 4-6. Although the Office of Technology Assessment report displays a high degree of concern for false positives and for the possibility of countermeasures introduced by those taking the test, in the trial context these problems are at least as severe when the examiner is a judge or juror and the witness is even a halfway skilled liar. *See* Zuckerman, DePaulo & Rosenthal, *infra* note 78.

72. One argument against reliance on polygraph procedures, even at apparently high levels of current accuracy, is that those who wish to deceive can use biofeedback to learn to control their bodily reactions and thereby to fool the test. L. SAXE, D. DOUGHERTY & T. CROSS, *supra* note 68, at 89. That, apparently, is what has already happened with demeanor evidence, though in more natural and less technologically self-conscious ways.

# TABLE 3A

## OTA POLYGRAPH DATA

Outcomes of Control Question Analog Studies of Validity

| | Guilty | | | | Innocent | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Number of subjects | Correct | Incorrect (False negative) | Inconclusive | Number of subjects | Correct | Incorrect (False positive) | Inconclusive | Total number of subjects | Lambda[k] |
| Barland and Raskin (21) | 36 | 63.9% | 8.3% | 27.8% | 36 | 41.7% | 16.7% | 41.7% | 72 | 0.47 |
| Podlesny and Raskin (127) | 20 | 69.0 | 16.0 | 15.0 | 20 | 91.0 | 4.0 | 5.0 | 40 | 0.75 |
| Raskin and Hare (137) | 24 | 87.5 | 0 | 12.5 | 24 | 75.0 | 4.0 | 20.8 | 48 | 0.83 |
| Rovner, et al.[a] (143) | 36 | 77.8 | 8.3 | 13.9 | 36 | 80.5 | 13.9 | 5.5 | 72 | 0.72 |
| Kircher (89a) | 50 | 60.0 | 4.0 | 36.0 | 50 | 76.0 | 2.0 | 22.0 | 100 | 0.72 |
| Dawson (49) | 12 | 91.7 | 0 | 8.3 | 12 | 58.3 | 25.0 | 16.7 | 24 | 0.33 |
| Widacki and Horvath[b,l] (207) | 20 | 90.0 | 5.0 | 5.0 | — | — | — | — | 20 | —[k] |
| Bradley and Janisse (35) | 96 | — | — | — | 96 | — | — | — | 192 | 0.33 |
| EDR | | 60.4 | 13.5 | 26.0 | — | 58.3 | 9.4 | 32.3 | | |
| Heart rate | | 35.4 | 20.8 | 43.8 | — | 33.3 | 19.8 | 46.9 | | |
| Szucko and Kleinmuntz[c] (160) | 15 | 71.3 | 28.7 | [d] | 15 | 49.3 | 50.7 | [d] | 30 | 0.22 |
| Ginton, et al. (62) | 2 | 100.0 | 0 | 0 | 13 | 84.6 | 15.4 | 0 | 15 | 0.00[l] |
| Honts and Hodes (75)[e]: | | | | | | | | | | |
| No countermeasurses | 9 | 67.0 | 0.0 | 33.0 | 12 | 33.0 | 17.0 | 50.0 | | 0.25 |
| Countermeasures | 19[f] | 58.0 | 5.5 | 36.6 | —[g] | — | — | — | | — |
| | 28 | | | | 12 | | | | 40 | |
| Honts and Hodes (76): | | | | | | | | | | |
| No countermeasures | 19 | 84.2 | 0.0 | 15.8 | 19 | 31.6 | 15.8 | 52.6 | | 0.53 |
| Countermeasures | 19 | 36.8 | 26.3 | 36.8 | —[g] | — | — | — | | — |
| | 38 | | | | 19 | | | | 57 | |
| Heckel, et al. (74)[h]: | | | | | | | | | | |
| Normals | —[h] | | | | 5 | 100.0 | 0.0 | 0.0 | | —[k] |
| Nondelusional psychiatric | —[h] | | | | 5 | 70.0 | 10.0 | 20.0 | | |
| Delusional psychiatric | —[h] | | | | 5 | 45.0 | 35.0 | 20.0 | 15 | |
| Hammond (74a) | 32 | 71.9 | 3.0 | 25.0 | 30 | 40.0 | 20.0 | 40.0 | 62 | |

[a] Summed across conditions.
[b] Examiner's task was to detect the one guilty person in each of 20 groups of four suspects.
[c] Based on ratings of 5' on a 1 to 8 scale of certainty of nondeception/deception.
[d] Examiners were not allowed to categorize an examination as inconclusive.
[e] Original subject assignments, 12 to each of 4 (including 2 countermeasure) conditions. A total of five countermeasure subjects were eliminated from the analysis of results for guilty subjects for failure to follow countermeasure instructions. Three no countermeasure subjects were eliminated for spontaneously using countermeasures.
[f] Nine guilty subjects used pain countermeasures (tongue biting) and 10 used a muscle (toe pressing) countermeasure.
[g] Innocent subjects used no countermeasures.
[h] There was no guilty condition.
[i] Not included in analysis reported in table 8.
[j] Lambda is a poor statistic when the base rate is skewed.
[k] Lamba was not calculated when only guilty or innocent subjects were used.

Source: OFFICE OF TECHNOLOGY ASSESSMENT, THE SCIENTIFIC VALIDITY OF POLYGRAPH TESTING: A RESEARCH REVIEW & EVALUATION 63 (1983).

# TABLE 3B

## OTA POLYGRAPH DATA

Outcomes of Field Studies of Validity

| | Guilty | | | | Innocent | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Number of cases | Correct | Incorrect (false negative) | Inconclusive | Number of cases | Correct | Incorrect (false positive) | Inconclusive | Total number of cases | Lambda |
| Bersh (29) (panel of 4) GOT unanimous | 32 | 96.9% | 3.1% | 0[a] | 36 | 88.9% | 11.1% | 0[a] | 68 | |
| ZOC unanimous | 38 | 89.5 | 10.5 | 0 | 51 | 94.1 | 5.9 | 0 | 89 | |
| Average unanimous | 70 | 93.2 | 6.8 | 0 | 87 | 91.5 | 8.5 | 0 | 157 | 0.84 |
| Majority (ZOC and GOT) | 34 | 70.6 | 29.4 | 0 | 25 | 80.0 | 20.0 | 0 | 59 | 0.82 |
| | | 81.9 | 18.1 | 0 | | 85.6 | 14.3 | 0 | 216 | |
| Horvath and Reid (84) (1 examiner, 10 examiner/evaluators) | 20 | 85.0 | 15.0 | 0[b] | 20 | 90.5 | 9.5 | 0[b] | 40 | 0.76 |
| Hunter and Ash (85)[c]—(1 examiner, 7 examiners/evaluators) | 10 | 87.1 | 11.4 | 1.4 | 10 | 86.4 | 14.1 | 0 | 20 | 0.74 |
| Skowick and Buckley, (155)[d]—(random selection; 7 examiner/evaluators) | 15 | 84.0 | 15.3 | 0.7 | 15 | 90.7 | 6.6 | 2.7 | 30 | 0.77 |
| Wicklander and Hunter (205)—(2 examiners/6 evaluators) | | | | | | | | | | |
| PG+[e] | 10 | 98.6 | 1.3 | 0 | 10 | 86.6 | 8.3 | 5.0 | 20 | |
| PG | | 90.0 | 8.3 | 1.6 | | 86.6 | 5.0 | 8.3 | | |
| Average | | 94.4 | 5.0 | 1.0 | | 86.6 | 6.7 | 6.7 | | 0.88 |
| Horvath (82)[f] (10 examiner/evaluators) | | | | | | | | | | |
| Verified cases | 28 | 77.1 | 22.9 | 0[g] | 28 | 51.1 | 48.9 | 0[g] | 56 | 0.28 |
| Davidson (47)[h]—(random selection 7 examiners/evaluators) | 10 | 90.0 | 10.0 | 0 | 11 | 91.0 | 0 | 9.0 | 21 | 0.90 |
| Raskin,[d,i] | 12 | | | | 4 | | | | 16 | |
| (1 examiner, 25 evaluators) Numerical | | 91.7 | 0 | 8.3 | | 75.0 | 0 | 25.0 | | 0.75 |
| Nonnumerical | | 83.3 | 8.3 | 8.3 | | 25.0 | 50.0 | 25.0 | | |
| Barland and Raskin (22) (1 examiner, panel of 5) | | | | | | | | | | |
| Panel[j] | 47 | 91.5 | 0 | 8.5 | 17 | 29.4 | 52.9 | 17.6 | 64 | 0.29 |
| Judicial outcome | 33 | 90.9 | 0 | 9.1 | 8 | 12.5 | 75.0 | 12.5 | 41 | 0.13 |
| Klemmuntz and Szucko (92)—(6 examiners/evaluators)[k] | 50 | 75.0 | 25.0 | 0[l] | 50 | 63.0 | 37.0 | 0[l] | 100[b] | 0.38 |

[a] Data for Inconclusives not reported; Inconclusives appear to total 27 (243 initial N—216 decisions reported).
[b] Examiner/evaluators were not allowed to judge charts as inconclusive as to overall deceptiveness. In another type of analysis on a question by question basis, judgments of doubtful or inconclusive were allowed.
[c] Average of two blind chart analyses spaced at least 3 months apart, done by same examiners.
[d] Average inquiries divided by number of examiners.
[e] PG+ indicated evaluators had access to written information in addition to polygraph charts (e.g. case details, subject behavior during examinations, etc.). Both PG only and PG+ examinations were done by the same examiners 2 months apart.
[f] Excludes Horvath's analysis of 26 unverified cases; because there is no criterion, this is a reliability study.
[g] There was apparently 15 (1.3 percent) inconclusive judgments out of 1,120 total judgments (10 examiners x 112 cases) which the author excluded from further analyses.
[h] Majority decision only.
[i] Seven examiners used numerical scoring; 18 used nonnumerical scoring procedures.
[j] Excludes 28 cases for which the panel was unable to come to a decision as to guilt or innocence.
[k] Decisions were based on one polygraph chart; standard practice generally employs at least three. Also, the evaluations were made by students with little polygraph experience.
[l] Inconclusives were not allowed.

NOTE: GOT = general question test.
ZOC = Zone of comparison.

SOURCE: OFFICE OF TECHNOLOGY ASSESSMENT, THE SCIENTIFIC VALIDITY OF POLYGRAPH TESTING: A RESEARCH REVIEW & EVALUATION 53 (1983).

the admission of polygraph examination evidence and rests heavily on demeanor evidence and the factfinder's assumed ability to use it. Where polygraph evidence is admissible at all, the admission is almost always made by stipulation between the parties.

Courts never have said, and do not now say, that they will never admit polygraph evidence. They have said,[73] and continue to say,[74] that if the test ever achieves an acceptable level of accuracy, rational courts will welcome it. But consider these generous views of the Arizona Supreme Court in *State v. Valdez*, and ask what they may reveal about the law's underlying attitude toward polygraph testing:

> A conservative estimate of the accuracy of such tests is as follows
> (1) In 75-80 per cent of the cases the examination correctly indicates the guilt or innocence of the accused;
> (2) in 15-20 per cent of the cases the results are too indefinite to warrant a conclusion by the examiner one way or the other;
> (3) 5 per cent or less is the margin of proven error.
>
> With improvement in and standardization of instrumentation, technique and examiner qualifications the margin of proven error is certain to shrink . . . . Although much remains to be done to perfect the lie-detector as a means of determining credibility we think it has been developed to a state in which its results are probative enough to warrant admissibility *upon stipulation*.[75]

Assuming the court believes its own conclusions about the data, why does the court think that 5 percent error leaves a lot of room for improvement? How much other evidence that factfinders receive has as little as 5 percent error? And if 5 percent error is the stage at which evidence can be admitted *upon stipulation*, how accurate does evidence have to be simply to be admitted? Perhaps the usual rules of admissibility do not apply when it comes to polygraph testing (and certain other kinds of evidence). Or perhaps the problem is not the perceived level of inaccuracy but rather the perceived level of accuracy.[76]

## F. Demeanor Evidence

The law cuts a large window for factfinders to observe a witness and determine whether he is telling the truth. The law urges factfinders to observe the demeanor of witnesses and draw from those observations whatever inferences they may about the witnesses' credibility. Here is what may be the original polygraph: the factfinder's eyes and ears and his intuitions about the information they provide.

---

73. Frye v. United States, 54 App. D.C. 46, 47, 293 F. 1013, 1014 (1923).

74. Witherspoon v. Superior Court of Cal., 133 Cal. App. 3d 24, 28, 183 Cal. Rptr. 615, 617 (1982).

75. State v. Valdez, 91 Ariz. 274, 282-83, 371 P.2d 894, 900 (1962) (emphasis added).

76. In authorizing the admissibility of low validity predictions of dangerousness in hearings to determine whether the death penalty should be imposed, the U.S. Supreme Court approved such expertise, stating that psychiatrists are not "always wrong with respect to future dangerousness, only most of the time." If a very low level of accuracy is sufficient for admissibilty on important questions of life and death, then perhaps it is high levels of accuracy that present a problem. That is to say, perhaps the more accurate something gets, the more objectionable it may be.

The law appears quite certain that jurors have the ability to make these observations and to draw useful inferences from them. Accordingly, it protects jurors' freedom and power to use such information. Learned Hand summarized this view:

> The words used are by no means all that we rely on in making up our minds about the truth of a question that arises in our ordinary affairs, and it is abundantly settled that a jury is as little confined to them as we are. They may, and indeed they should, take into consideration the whole nexus of sense impressions which they get from a witness. This we have again and again declared, and have rested our affirmance of findings of fact of a judge, or of a jury, on the hypothesis that this part of the evidence may have turned the scale.[77]

Given the law's reliance on demeanor evidence, one might think that social science research on the detection of deception through ordinary observation would cast considerable light on an important assumption. Fortunately, a genre of psychological research is concerned with non-verbal behavior and the cues it provides for detecting truth-telling or lying. Zuckerman, DePaulo, and Rosenthal[78] have carried out an extensive meta-analysis[79] of the non-verbal research literature on this topic.

Several important conclusions flow from their data. Most important, perhaps, is that people are not good at sensing whether someone is telling the truth.[80] Decisions about whether a statement is the truth or a lie are made about as well as if one were tossing a coin. Second, the problem is not simply that certain non-verbal patterns do not correlate with lie-telling; the patterns upon which observers rely differ from the patterns that actually occur. Observers correctly believe that people are more likely to be lying if they smile less, commit more speech errors, speak more hesitantly, and speak in a higher pitch. At the same time, however, they erroneously assume that lying is signaled by less gazing, more postural shifting, longer speech latencies, and slower speech rate. Furthermore, observers make no use of some behavior which *does* signal lying: appearing less involved in their communication, appearing less spontaneous, showing negative affect, making more negative verbal statements, and using more adaptors.

Third, it is interesting to note that the unschooled law, like the unaided and untrained observer, has been fooled. Table 4 summarizes the findings of a number of studies on types of evidence that improve or diminish a person's ability to detect lying. Visual access to the speaker's face *reduces* the accuracy of detection of deception compared with the same modes of observation excluding facial cues (1.49 to 1.00). An audiotape of the testimony with no access to demeanor at all is better (1.09 to 1.00). A written transcript is not

---

77. Dyer v. MacDougall, 201 F.2d 265, 269 (2d Cir. 1952).

78. Zuckerman, DePaulo & Rosenthal, *Verbal and Nonverbal Communication of Deception*, in 14 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 1 (L. Berkowitz ed. 1981).

79. A meta-analysis is a statistical aggregation and analysis of a body of data contained in separate studies. For explanations of such combinational research methods, see generally Rosenthal, *Combining the Results of Independent Studies*, 85 PSYCHOLOGICAL BULL. 185 (1978); R. LIGHT & D. PILLEMER, SUMMING UP: THE SCIENCE OF REVIEWING RESEARCH (1984).

80. Zuckerman, DePaulo & Rosenthal, *supra* note 78, at 39.

dramatically worse than speech plus demeanor (.70 v. 1.00).[81] The best performance is obtained when observers have visual access to body movements (but not the face) plus the sound of the voice (1.49). Of course, even the best of these still is not a great deal better than coin tossing.

TABLE 4

META-ANALYSIS OF NON-VERBAL DETECTION OF DECEPTION

ACCURACY OF DETECTING DECEPTION (IN STANDARD DEVIATION UNITS)

| | Visual cues | | | | |
|---|---|---|---|---|---|
| | Face | | No face | | |
| Auditory cues | Body | No body | Body | No body | Means |
| Speech | 1.00(21)[a] | .99(9) | 1.49(3) | 1.09(12) | 1.14 |
| No speech | .35(6) | .05(7) | .43(4) | .00[b] | .21 |
| Means | .68 | .52 | .96 | .54 | .68 |

Transcript only: .70(6)
Tone only: .20(4)

[a] Number of studies upon which *d* is based is enclosed in parentheses.
[b] Theoretical accuracy.
Source: Zuckerman, DePaulo & Rosenthal, *Verbal and Nonverbal Communication of Deception*, 14 ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY 26 (L. Berkowitz ed. 1981).

The explanation for these phenomena appears simple enough: Observers rely on faces to give important signals, and because liars think the same thing, they control their facial movements. The result is that observers are distracted and miss the most revealing cues. Contrary to its assumptions, the law, therefore, might improve the ability of jurors to assess credibility by covering the faces of witnesses with masks.

## G. Eyewitness Accuracy

The question of the accuracy of eyewitness identification has occupied a huge amount of the attention of experimental and social psychologists interested in making contributions to the legal system. Findings in this area could aid courts in setting psychometrically informed standards for out-of-court identifications (lineups, photospreads) and other observations, and could inform factfinders whether the circumstances surrounding the observation were such as to enhance or degrade the likelihood of an accurate report by the witness.[82] The reaction of the courts to such learning has been

81. *Id.* at 27. Consider what this research does to the judicial assumption that appellate judges are in no position to assess the credibility of testimony—which they traditionally receive only in written form—compared to jurors or trial judges, who had an opportunity to observe the demeanor of the testifying witness.

82. For analyses of eyewitness accuracy, see generally E. LOFTUS, EYEWITNESS TESTIMONY (1979); A. YARMEY, THE PSYCHOLOGY OF EYEWITNESS TESTIMONY (1979); Buckhout, *Eyewitness Testimony*, SCI. AM., Dec. 1974, at 23; Buckhout & Greenwald, *Witness Psychology*, in SCIENTIFIC & EXPERT EVIDENCE 1291 (E. Imwinkelried ed. 1981).

mixed, but for the most part they have held such information to be superfluous (jurors already know about the limitations on eyewitness perception, memory, and retrieval) or to invade the province of the jury (the data might seem to tell the jurors how much weight they ought to give to a particular eyewitness' testimony).[83] Courts might want to weigh the importance of avoiding error in deciding how much accuracy is needed and how much time to devote to achieving greater accuracy.[84]

## H. Instructions

Several different lines of research have all concluded that the instructions that judges typically give to juries are confusing at best, and sometimes downright incomprehensible.[85] These studies also demonstrate that by using a number of techniques for improving comprehensibility, courts could improve instructions and make them easier to understand, remember, and apply.

In light of these findings, then, the following question arises: If the instructions that judges give to juries are the essential link for transmitting to the decisionmakers the rules of decision that courts and legislatures have arduously worked out to their present state of development, and if the possibility exists for more effective transmission of that information to jurors, then why have courts resisted making improvements in the process of instructing jurors? To forego this vital link is to keep the world of statutes and appellate opinions from influencing jury decisionmaking. It is to render the work of legislatures and appellate courts more or less irrelevant, and to invite—indeed, to require—jurors to borrow rules of decision from the general culture or to make up their own as they go along.[86]

Additional studies pertain to the effects of jury instructions to use evidence in some limited way or to disregard entirely inadmissible evidence that has inadvertently entered the trial. These studies variously show that such instructions have no effect or have an effect contrary to the one intended.[87]

---

83. However, the California Supreme Court has held that in appropriate cases the exclusion of expert testimony on eyewitness accuracy is reversible error. People v. McDonald, 37 Cal. 3d 351, 377, 208 Cal. Rptr. 236, 254, 690 P.2d 709, 727 (1984). Most jurisdictions leave the matter to the discretion of the trial judge, and they are uneven about admitting or excluding. Although this issue usually arises in criminal cases, eyewitness testimony is familiar in some kinds of civil suits, especially torts.

84. Or adopt certain research findings as a matter of law and give them as a jury instruction. *See* Walker & Monahan, *supra* note 48.

85. *See generally* Charrow & Charrow, *Making Legal Language Understandable: A Psycholinguistic Study of Jury Instructions*, 79 COLUM. L. REV. 1306, 1358-59 (1979); Severence & Loftus, *Improving the Ability of Jurors to Comprehend and Apply Criminal Jury Instructions*, 17 LAW & SOC'Y REV. 153, 153-55 (1982); Sales, Elwork & Alfini, *Improving Comprehension for Jury Instructions*, in PERSPECTIVES IN LAW & PSYCHOLOGY 23 (B. Sales ed. 1977).

86. The Model Health Care Provider Liability Reform Act recommends adoption of "plain, simple and readily understandable" language in jury instructions and the comprehensive use of interrogatories to reduce unwanted variation in jury decisions. Model Health Care Provider Liability Reform Act § 13 (1988).

87. Broeder, *The University of Chicago Jury Project*, 38 NEB. L. REV. 744, 754 (1959); Wegner, Schneider, Carter & White, *Paradoxical Effects of Thought Suppression*, 53 J. PERSONALITY & SOC.

For example, in a mock jury experiment, when jurors knew nothing about the defendant's insurance coverage, the average award to the plaintiff was $33,000. When jurors learned that the defendant did have coverage, the average award grew to $37,000. But when jurors learned of the coverage and then were admonished to disregard that fact, the average award rose to $46,000.[88] The practical reform implied—to ignore rather than draw attention to such inadmissible evidence—creates a dilemma for the courts. Is it preferable to do something practical about the problem of biasing the factfinder, or to make a symbolic statement that demonstrates the law's concern but actually makes matters worse? If this study's findings are generally true, the law has an important choice to make; we cannot have it both ways.

## I. Larger System Design

Perhaps the grandest feature of a legal system is the fundamental strategy by which it strikes out in pursuit of decisions. The Anglo-American system is, of course, directed at nearly every turn by its adversarial strategy. The major alternative to this adversarial type of system would be an inquisitorial one. The line of research begun by Thibaut and Walker,[89] known as "procedural justice," provides the first systematic empirical account of the difference in effect of these two alternative approaches to designing a trial system. Among other findings, Thibaut and Walker discovered that the adversarial system made people feel more satisfied with the outcome of a judicial proceeding.[90] This was true for uninvolved observers as well as for the parties, and for people living in continental nations as well as in the United States.[91] And while the original research was based on simulations of business problems and trials, it has been extended by later research into more natural settings, and the findings have been confirmed.[92]

Subjective satisfaction of participants and observers is only one of the dependent variables that the procedural justice researchers have examined. Another interesting variable is the distribution of facts (those favorable to the plaintiff versus those favorable to the defendant) delivered to the trial under the two schemes. Thibaut and Walker found that the inquisitorial system tended to deliver to the factfinder an image of the facts that reproduced the underlying distribution of facts that existed in the case. The adversarial system tended to produce a more balanced array of facts from an initially skewed distribution. That is to say, the adversarial system tends to make a dispute into a more equal fight even when the underlying fact distribution

---

PSYCHOLOGY 5 (1987); Wissler & Saks, *On the Inefficacy of Limiting Instructions*, 9 LAW & HUM. BEHAV. 37, 46 (1985).

88. Broeder, *supra* note 87, at 754.

89. J. THIBAUT & L. WALKER, *supra* note 8.

90. *Id.* at 77.

91. *Id.* at 77-80.

92. Such work is reviewed in E. LIND & T. TYLER, THE SOCIAL PSYCHOLOGY OF PROCEDURAL JUSTICE 211-14 (1988).

favors one side. That balancing is produced, presumably, by advocates who are motivated to seek out obscure evidence, to invent favorable interpretations of unfavorable evidence, and to argue for their client's preferred conclusions.

This finding presents would-be designers of a legal system with an interesting dilemma: Should they aim for a system that reproduces facts more accurately, or one that promotes greater citizen satisfaction and support?[93] For present purposes, I simply want to draw attention to the finding that the adversarial system has the power to bring out of clarity a potentially useful ambiguity.

## III

## THE RESTRAINT OF ACCURACY

The examples of social-psychological research on the justice system given in the preceding section do not, of course, exhaust the supply, but they do illustrate this genre of research and its potential benefits. Empirical social scientists approach problems on the assumption that solutions can be found—or at least that there are better and poorer choices to be made. If one can identify certain values to be maximized, an empirical comparison of alternatives can show which alternatives yield the most desirable results in light of those values.[94] Social and experimental psychology have knowledge and methods to bring to the law to help determine which structures and procedures do a more accurate job of finding facts and minimizing error and unpredictability.

Among these examples I have tried to emphasize findings bearing on potential improvements that the law has not adopted. More interestingly, I have included examples of the obverse, that is, harmful changes which the law has adopted. This is, of course, an unfair selection of evidence. But a systematic appraisal of the degree to which the law has responded—positively or negatively, purposefully or fortuitously—to all social science empirical findings would be a massive and perhaps impossible undertaking. I seek only to establish that certain seemingly irrational categories of the law's responses to such knowledge exist.

---

93. Thibaut & Walker, *supra* note 7, argue that, considering that most of the disputes presented to the legal system are distributional, or "justice," conflicts, resolution through finding the truth is not an effective strategy. Just allocations in the face of competing claims are best achieved by the pursuit of satisfying outcomes; there are no "true" outcomes. *Id.* at 548-49. (In such cases, facts are at best the supporting cast; they play no starring role. *See* M. SAKS & R. VAN DUIZEND, THE USE OF SCIENTIFIC EVIDENCE IN LITIGATION 6 (1983).) By contrast, purely or largely cognitive conflicts—the sorts of problems that scientists confront—are not likely to be best solved by adversary procedures. Thibaut & Walker, *supra*, at 556.

94. Evaluation research is the operational expression of this. The literature on the subject is vast. For a publication devoted entirely to this field, see the collected volumes of EVALUATION STUDIES REVIEW ANNUAL.

## A. The Flight from Apparent Accuracy

Some of these examples illustrate the law refusing to make use of help that even some courts have acknowledged is available (for example, polygraph examinations), while still relying on older forms of help that are demonstrably poorer (for example, demeanor evidence). Reasons have been suggested earlier. Some courts think that the technique has not yet achieved the requisite level of validity, but the *Valdez* opinion shows that explanation to be paradoxical at best. Another commonly offered explanation is that credibility assessment is the factfinder's business, and the law is loathe to let anyone else assist the factfinder in this area. Thus, error-laden credibility assessment by judges and jurors is preferable to more accurate assessments by polygraph examiners. Under this view, even if examiners achieved perfect accuracy, their evidence could not be provided to the factfinder.

Another example of an apparent flight from accuracy is the use of probability data. The courts' reaction to the use of such data as primary evidence (for example, persons with a certain exposure to a certain carcinogen are $x$ times more likely to develop cancer than unexposed persons; a defendant with blood group type $Z$ is $x$ times more likely to be the father of the child than a person chosen at random from the population) has been highly inconsistent, as we have discussed. The reasons given by courts when excluding or when admitting probability evidence are merely truisms about this kind of evidence: general probabilities used to resolve particular uncertainties. What is said about the evidence admitted is also true of the evidence excluded; what is said about the evidence excluded is also true about the evidence admitted.

Stating countervailing truisms does not begin to respond to the question of when such data are to be admissible (or sufficient to satisfy the burden of proof) and when they are not.[95] Neither courts nor scholars have been able to provide an answer to this question.[96]

One answer may be that there is no answer: Such evidence always is both probative and yet ultimately incomplete or imperfect. Courts may choose to reject probability evidence. In this case they choose to limit factfinders to less complete or less perfect evidence for none of the usual policy reasons.[97] On the other hand, courts may choose to admit probability evidence when filtered through a clinical expert but not when presented through a scientific expert in

---

95. The likely exception to this situation is in class actions, where the probability data can be about the class itself, and thereby provide a level of data (aggregate) that equals the parties (aggregate). But when it comes to damages, though on average the plaintiff class and the defendants might be fairly treated vis-à-vis each other, as among the plaintiffs (or among several defendants) some will be overcompensated and others undercompensated (or assessed too much or too little). By the nature of these data, we can never know who has won or lost too little or too much.

96. A possible answer is suggested by the structure of Chapter 4 of J. MONAHAN & L. WALKER, *supra* note 9, at 159-274. That chapter divides the use of empirical social science data by the law into future, present, and past factfinding. (Judges seem most willing to use such information for establishing the identity of future actors and least willing to use it to establish past acts.) *Id.* at 228.

97. For the general policy of admitting all relevant evidence, see generally FED. R. EVID. art. IV.

a more logical (and more explicit) trial sequence.[98] In this case, the courts are unwittingly choosing to (a) keep the data implicit rather than explicit, less subject to full examination; (b) perhaps keep it from the factfinder altogether; and (c) transfer the decisionmaking responsibility that would otherwise be the factfinder's to the witness. Surely this creates no increased expectation that the factfinder will reach the correct answer. Why do courts "not honor their own formulas"[99] when dealing with such evidence?

When probability data are offered to provide baserates (especially evidence on the evidence) or when other kinds of evidence are offered to help jurors assess primary witness evidence (such as expert testimony on eyewitness accuracy), the probability data really amount to another layer of evidence: second order knowledge. The primary witness takes the stand to say *x* and the secondary knowledge witness takes the stand to tell the jury something about the first witness. This process raises concerns about raising the costs of the search for truth beyond a level that the dispute resolution enterprise can tolerate. Those concerns may be most valid with respect to experts on eyewitnesses, who then are countered by other contrary experts, sending the trial off on what may be a detour that in the end provides little or no improvement in factfinding. However, this reasoning may not suffice as an explanation of the wholesale rejection of second order knowledge for two reasons. First, such evidence is a kind of knowledge that could be judicially

---

98. What I have in mind here is illustrated by the following alternative patterns for presenting the same information. The information consists of: (1) general scientific facts concerning the phenomenon that is to inform the factfinding (for example, that symptoms *a*, *b*, and *c* usually are associated with condition *X*); and (2) particular clinical facts concerning the case at hand (the plaintiff has symptoms *a*, *b*, and *c*).

Pattern A: The clinical expert (physician, psychiatrist, auto mechanic, etc.) presents the opinion that the plaintiff has condition *X*. Under further examination the witness may or may not be asked to explain the major and minor premises that led the witness to the conclusion. The jury is only sure to hear the expert's conclusion. The jury's choice is whether to believe the expert's conclusion.

Pattern B: A scientific expert is offered to present the general scientific facts, the major premise, that symptoms *a*, *b*, and *c* usually are associated with condition *X*. Then a clinical expert is presented to testify about the extent to which symptoms *a*, *b*, and *c* were observed in the plaintiff. The jury's responsibility, then, is to draw the conclusion as to whether the plaintiff has condition *X*.

Under Pattern B, the factfinder gets the relevant information, weighs the information, and draws its own conclusions. Under Pattern A, the witness makes the decision (about condition *X*) for the factfinder and, if the witness is not asked to reveal his underlying observations and reasoning, the factfinder is left in the position of drawing conclusions based on the witness' credibility—probably derived from superficial characteristics of the witness rather than the substance of the witness' knowledge.

Pattern B may be the more logical. But Pattern A is the one to which the law is more accustomed. Florida v. Zamora (Cir. Ct. Dade County, Fla. 1977), Trial Transcript, Case No. 77-2566, *reprinted in* J. MONAHAN & L. WALKER, *supra* note 9, at 313, provides an illustration of a case in which a lawyer was too determined to present evidence in the order of Pattern B. The presiding judge was perhaps overly comfortable with Pattern A and insisted on receiving expert testimony in that pattern or not at all.

For information on another aspect of this general problem, see Monahan and Wexler's discussion of the need to unpack the sometimes conclusory and imperialistic statements of expert witnesses in order to see when and where they have overstepped the bounds of their expertise and power. Monahan & Wexler, *A Definite Maybe: Proof and Probability in Civil Commitment*, 2 LAW & HUM. BEHAV. 37 (1978).

99. H. HART & J. MCNAUGHTON, *supra* note 51.

noticed and given by judicial instruction. If it were shown that it added much accuracy at little cost, there would not be much reason for eschewing it. As it is now, jurors are asked to argue about it among themselves based on their various experiences with the problem. Second, some kinds of second order knowledge would result in certain kinds of experts being barred from testifying. Perhaps, for example, document examiners would no longer be allowed to testify about the source of handwriting. Such a bar might result in a widespread savings of judicial resources with no significant decrease in accuracy. Again, the flight from improved accuracy needs more of an explanation than has been offered.

### B. The Pursuit of Apparent Inaccuracy

In a sense, anything that appears in the preceding subsection has its counterpart in this subsection. Every time the law rejects information that would enhance factfinding it is at some level choosing to reduce the likelihood of an accurate result. In this subsection I focus on choices that amount to something more than the trend toward inaccuracy as a mere consequence of a series of flights from accuracy.[100] The examples below are somewhat more independent, if not affirmative, steps at restraining accuracy by actively pursuing inaccuracy.

Examples of this phenomenon include the reduction in jury size, the use of incomprehensible instructions, and the adversary system itself as a process for factfinding. Whatever else may commend it to us, accurate factfinding is not one of the advantages of the adversary system of dispute resolution.[101] The choice of this fundamental procedural structure as the means of organizing our litigation process may simply indicate that accurate factfinding is not the dominant value or purpose of the legal process, and that we are willing to give up some accuracy in factfinding in order to gain something else. The reduction of jury size (when the traditional larger juries produce more consistent and stable trial results) and the use of incomprehensible instructions (when more lucid and coherent ones are possible and would make the jury's decisions reflect legal rules rather than idiosyncratic criteria) are much more of a puzzle.

Notwithstanding findings regarding reduction of jury size and decreased accuracy, courts and legislatures in recent decades have approved smaller

---

100. None of this is to say that any of these by-products are not serious. A court that prevents a jury from having its faith in ballisticians or toxicologists or eyewitnesses shaken may be actively engaged in insulating these kinds of witnesses from attack. Some scholars have termed this approach to factfinding "the exorcism of ignorance"—a third alternative to the model of a trial as a "search for truth" or a "fair fight." *See* Risinger, Denbeaux & Saks, *supra* note 41, at 49-52. According to this view, even if we cannot replace our ignorance with understanding, we can place our trust in experts who will tell us what to believe, even though we have no demonstrable basis for believing that they know what they are talking about. To allow data that tend to announce that the emperor-expert has no clothes would vitiate this function of the trial. *Id.*

101. Thibaut & Walker, *supra* note 7.

juries.[102] Thus, they have moved in the direction of *increasing* the error and unpredictability of the decisions of civil juries. This increase applies to both the decision as to liability and the size of damage awards.[103] What makes the reduction of jury size a particularly interesting example is not only the clarity with which this legal policy change increases error variation, but also the ease with which the law could fine tune the level of error by adjusting the jury size.

Perhaps when increasing error becomes intolerably troublesome, the law can readjust the level of error by once more changing[104] the jury size (upward). Thus, the U.S. Department of Health and Human Services ("DHHS") recently recommended in its Model Health Care Provider Liability Reform Act that states require juries of twelve in malpractice cases for the express purpose of increasing stability and predictability in trial outcomes.[105]

The DHHS lawyers regard the absence of such stability as contributing to greater transaction costs of litigation (if settlement becomes more difficult and more cases go to trial) and greater insurance costs (if the insurers' response to more variable awards is to prepare for the worst in all cases).

Generally, the cost of avoiding variability and inaccuracy by adjusting jury size and jury instructions would be little, and the apparent benefits would be great. Therefore, the law's choices to avoid these advantages truly call for some explanation, lest we conclude the law sometimes unreasonably prefers the irrational over the rational.

## IV

## WHY DOES THE LAW MAKE SUCH CHOICES?

I have not argued or implied that the law is engaged in some perverse search for wrong answers. Surely the law has taken many steps to structure its factfinding process in ways that seem calculated to increase the ability of the parties to present relevant and probative evidence to factfinders. The rules of procedure[106] and evidence[107] are filled with examples of efforts to enhance the capacity of the factfinding process to obtain correct answers. Indeed, the first data I presented in this article suggest that the legal decisionmaking process achieves an impressive level of reliability.[108]

I have argued that the law *sometimes* rejects procedures or categories of evidence that offer clear potential to enhance the factfinding process, and that the law sometimes even pursues policies that seem sure to restrain the level of

---

102. The effects of sample size on standard error have been well understood since the end of the nineteenth century and the development of the central limit theorem. *See* J. SCHUTTE, EVERYTHING YOU ALWAYS WANTED TO KNOW ABOUT ELEMENTARY STATISTICS 109-10 (1977). The origins of this theorem date back to 1733, with the work of De Moivre. *Id.* at 192.

103. These effects were discussed in more detail in M. PETERSON & G. PRIEST, *supra* note 34, and the effect on damage awards is presented visually in Figure 1.

104. Or at least proposing to change.

105. Model Health Care Provider Liability Reform Act § 13(3) (1988). The rationale is given in the "Description" accompanying the Act.

106. FED. R. CIV. P. 26 (concerning discovery, for example).

107. FED. R. EVID. arts. IV, VIII (concerning relevance and hearsay, for example).

108. *See supra* notes 13-17 and accompanying text.

accuracy that the factfinders can achieve. In short, the law does not pursue a policy of continual enhancement of the factfinding process. Why not?

A. Three "Simple" Possibilities

1. *Ignorance.* The law may, of course, suffer from a collective institutional ignorance about the sorts of findings discussed in Part II. If this is the case, then as soon as the courts or legislatures become better informed, they will change policies to comport with the research findings. While this possibility surely explains some of the law's policy choices, I do not think it can explain certain persistent patterns of restraint. In a number of the illustrations offered, the courts stared directly at the data, understood them, and yet plainly acted contrary to their implications.[109]

2. *Inertia.* Another popular explanation would be that the law is slow to stir itself from its old ways. It is so comfortable with its traditional forms that it prefers familiar old errors to strange new corrections. While this theory may explain some refusals to change, I find it generally unpersuasive because the law does change itself with considerable regularity. And if we look at the particular examples discussed above, some of them represent abrupt and widespread change from traditional forms (for example, jury size reductions) and leaps into the unfamiliar and uncomfortable (for example, forensic science).

3. *Other Values.* A third possible explanation is that accuracy is not everything, and that reliable and valid factfinding has merely been trumped by other legitimate values that the law is pursuing. This possibility was discussed at some length in Part I. Some of these other values are explicit in the Federal Rules of Evidence themselves. Examples include: exclusion of some relevant evidence because of time restraints (Rule 403), inadmissibility of compromise offers (Rule 408, 409), and of subsequent remedial measures (Rule 407), so as not to discourage their being taken. Other values include the preservation of party control of case presentation and the passivity of factfinders.

More elaborate versions of this view find expression in the writings of Resnik,[110] Mashaw,[111] Leff,[112] and other scholars. For example, in rejecting the introduction of some kinds of probability evidence as a basis for decisionmaking, Tribe[113] has argued that the law has to pretend that its processes have produced correct results, even when we can document and quantify the uncertainty that is inherent in the factfinding process, because one of the law's functions is to make us feel good about the results we have

109. *See, e.g.*, Ballew v. Georgia, 435 U.S. 223 (1978) (reaffirming constitutionality of the six-person jury, and ignoring studies detailing problems associated with this reduction); State v. Valdez, 91 Ariz. 274, 371 P.2d 894 (1962).

110. Resnik, *Tiers*, 57 S. CAL. L. REV. 837, 1006-16 (1984).

111. J. MASHAW, DUE PROCESS IN THE ADMINISTRATIVE STATE (1985) (discussing dignitary values in administrative law).

112. Leff, *Law and*, 87 YALE L.J. 989 (1978).

113. Tribe, *supra* note 49.

produced.[114] I do not doubt that the law is pursuing multiple and sometimes contradictory goals, and must inevitably make tradeoffs among competing values. But I do have some trouble with the notion that the trial process ought to be an ignoring of uncertainty or an "exorcism of ignorance"[115]—that is, a process by which we comfort ourselves in decisionmaking by allowing so-called experts to make decisions for us. In other contexts, such as selling products or delivering services, we might simply call this fraud.[116] When the "other values" argument goes this far, I think it goes too far.[117]

As noted in Part I, by not heeding some of the findings presented in Part II, the law seems to serve none of these "other values," but instead only reduces accuracy and raises costs, as when it uses unvalidated but asserted expertise. Thus, in such cases, the law serves only an "exorcism of ignorance" function. Surely we expect the law to be doing something more.

### B. One Not So Simple Possibility: Optimal Gray

The law may have less visible but more essential goals, such as maintaining the basic functioning of the overall system, which in non-obvious ways affect choices concerning individual procedural or evidentiary rules. These goals may even require that a margin of real error be preserved in decisionmaking. Let us call these larger and more essential goals the system's meta-goals.

1. *The Law's Procedural Meta-Goals.* These meta-goals might run along the following lines: In order for common law courts to announce, refine, revise, and reverse the law, they must decide cases—preferably cases that will contribute the most to the law. The courts' pronouncements do not serve the needs only of cases that go to trial, but also of cases that are settled and of the many social interactions that never become litigation because they are conducted "in the shadow of the law."[118] Thus, the sampling of cases—the grist for the case-deciding mill, the stuff for which the substantive rules exist—may be an important procedural meta-goal.

---

114. [T]he fact that [some mistaken verdicts] must occur if trials are to be conducted at all need not undermine the effort, through the symbols of trial procedure, to express society's fundamental commitment to the protection of the defendant's rights as a person, as an end in himself. On the other hand, formulating an "acceptable" risk of error to which the trier is willing deliberately to subject the defendant would interfere seriously with this expressive role of the demand for certitude—however unattainable real certitude may be, and however clearly all may ultimately recognize its unattainability.

*Id.* at 1374.

115. Risinger, Denbeaux & Saks, *supra* note 41. Sometimes the law does not evaluate evidence rationally. Instead, it calls on experts to assert a conclusion which is largely insulated from challenge. This may be an exorcism-of-ignorance function of law, enabling courts to escape the burden of making especially difficult or painful decisions.

116. "All professions are a conspiracy against the laity." G.B. SHAW, THE DOCTOR'S DILEMMA, Act I (1906).

117. That is to say, the law may from time to time *be* an exercise in the exorcism of ignorance. But I do not believe that we can seriously argue that it *ought* to be one as opposed to a means of providing the factfinder with useful, if unpleasant, facts and tasks.

118. *See generally* Mnookin & Kornhauser, *Bargaining in the Shadow of the Law: The Case of Divorce*, 88 YALE L.J. 950 (1979).

2. *Mechanisms to Advance the Meta-Goals.* If the courts have too few cases in their sample, they will not have enough exposure to social problems to provide effective guidance through wise rules; if they are overwhelmed by caseloads, they cannot give them the attention necessary to develop thoughtful law.[119]

One device that might be helpful for insuring the right amount and kind of sampling is the maintenance of an aura of error and unpredictability. If the courts became ever more accurate and consistent, then attorneys would become ever more able to predict the outcome of a trial without the trouble of going to trial. If the outcome of trials could be predicted correctly, there would be no need for trials; settlements might become swift and certain. Without trials there would be no judicial sampling of disputes: no verdicts, no appeals, no further development of normative principles by courts, and no principled guidance for future disputes that might arise in changed social contexts.

On the other hand, if there were no consistency in judicial decisionmaking, then too few or no cases would go to trial. If the outcome of a trial were no more predictable than the tossing of a coin or the spinning of a roulette wheel, those would be cheaper and faster devices for reaching equally good results. If attorneys' skills in understanding and arguing facts and applying doctrines played no part in determining the outcome of cases, then attorneys would never need to expend effort in acquiring facts and researching law and thinking about how to present them to a court. Games of pure chance require no skill.[120]

Thus, in order to keep the entire system operating, some level of uncertainty is helpful. Parties then have enough knowledge of what a court would do if their cases were to come before it[121] and can plan and assess the likely outcomes of the cases. Yet parties also remain uncertain enough so that they are eager to maintain outcome control by reaching settlements without trial in the great majority of cases. This range of functional ambiguity might be termed "optimal gray"—a level of predictability that is neither too low nor too high.

In addition, trial factfinders must not be allowed to dispose of cases in a way that precludes appellate review. For purposes of keeping workloads manageable, appellate courts might be pleased when cases are so well handled at the trial level that no questions requiring appellate review arise. But for purposes of allowing the law to grow, trials cannot be allowed to be so

---

119. Consider the perfunctory treatment that courts give cases when they are overburdened. The development of the law suffers. W. Bablitch, Court Reform of 1977: The Wisconsin Supreme Court 10 Years Later (1988) (LL.M. Thesis, University of Virginia Law School).

120. This point can perhaps be made best by relating an anecdote from my college days. A certain wagering game involving dice became popular for a brief time. One day several of my friends came upon another friend who was sitting at a table by himself rolling dice. When they asked what he was doing he quipped, "I'm practicing for tonight's game."

121. "The prophecies of what the courts will do in fact, and nothing more pretentious, are what I mean by the law." O.W. HOLMES, *The Path of the Law*, in COLLECTED LEGAL PAPERS 167, 173 (1920).

dominated by the factfinding process that appellate courts have no power, in principle, to review them. To allow this type of factfinding domination would damage the dispute-sampling function (from the viewpoint of appellate lawmaking).

At the same time that the law must preserve the possibility of appellate review of cases so that the more general doctrines are well tended, certain cases arise that need to depart from those doctrines.[122] The law cannot afford (in terms of its available resources or its perceived legitimacy) to allow these exceptions to be made explicitly and as a matter of some distinguishing principles. Procedures must exist to allow courts to decide these cases without disturbing or being disturbed by the general doctrines.

Let us reexamine some of the findings discussed earlier, now from the viewpoint of serving the meta-goals and maintaining a zone of optimal ambiguity.

*a. Instructions.* How does the judicial distaste for giving understandable instructions serve the meta-goals? First, instructions serve the more important purpose of reannouncing rules of the law. In doing so, it is more important for the trial court to echo the appellate courts than to communicate clearly to the jury. The real audience for those instructions is thus the legal community, not the jury. Second, by keeping the jury only loosely tethered to the law, the courts can keep the level of predictability from becoming too high. Each new jury can have a somewhat different understanding of what the applicable rule of law is. Third, the jury is freer to render individualized justice in the case at hand. Its verdict is important to the parties; it is far less important to other people in other cases. By giving legally correct, even though incomprehensible instructions, trial courts accomplish all of these good things. The first point enhances the law's accuracy and predictability, the second restrains its accuracy and predictability, and the third does both: enhancing the rendering of particularized justice while restraining the predictability of the application of the law.

*b. Demeanor "evidence."* Flexibility in the factfinding process simultaneously protects litigants from the rigid application of the law and gives appellate courts an opportunity to reinforce doctrines while allowing departures from those doctrines to be made (by the jury) in a way that protects the law from charges of inconsistency. If judges had to carve out explicit exceptions with more frequency than they already do, the legitimacy of the judicial process would be highly questioned. But if juries do so because of their first-hand observations of witness demeanor that is quite a different

122. Examples include the "widow's law," releasing widows from promises to pay their deceased husbands' debts. For more information on this and other special exceptions to contract doctrine, see Kronman, *Paternalism and the Law of Contracts*, 92 YALE L.J. 763 (1983).

story. The doctrine can remain intact; but the departure from doctrine in individual cases is protected.[123]

Certain advances in accurate factfinding would interfere with these more important functions of adjudication. If assessment of demeanor were recognized as less diagnostic of lying than the reading of a transcript or viewing of a videotaped record of a trial, then the jury's departure from the correct result on the facts and the doctrine would lose its protective insulation. To disregard the research on detection of deception and to continue to act as though factfinders at trial have special powers to assess deception is to adopt a legal fiction that is helpful to the overall process. To incorporate such research might do more harm than good to the overall process. It does not matter if demeanor provides no cues or even if it provides miscues to credibility; the vague ideas of demeanor and credibility assessment serve a useful function. Social science research on the detection of deception shows that the emperor has no clothes, and incorporating that research into the law procedurally would diminish the meta-goal of optimal ambiguity even if it would lead to procedures that improved factfinding in the individual case.

*c. Polygraph.* The same is true of polygraph testing. The problem is not that polygraph examinations are worse than jurors; it is that they are or threaten to be better. And the better they become, the greater a hazard they will pose. Imagine that polygraph testing produced a 100 percent accurate evaluation concerning whether any given statement by a witness were true or false. The limitations that such polygraph evidence would place on factfinders would greatly hamstring the legal process. It would produce more predictable results, compel adherence to general doctrine in particular cases, upset the balance between more or less fixed doctrines tempered by individualized justice, and begin to interfere with the law's need to sample cases. In short, it might shift trial outcomes out of the zone of optimal gray.[124]

---

123. Consider the following cases in which juries, without disobeying the law, made improbable findings of fact, based in part on witness demeanor: Calvanese v. Babcock, 10 Mass. App. 726, 412 N.E.2d 385 (1980); Sears, Roebuck & Co. v. Roque, 414 N.E.2d 317 (Ind. Ct. App. 1980); Berry v. Chaplin, 74 Cal. 2d 652, 169 P.2d 442 (1946). In each of these cases, expert testimony was disregarded in favor of contrary lay testimony, despite the implausibility of the lay testimony. Arguably, the jury was creating new substantive rules, or at least carving out special exceptions to the law, in order to reach a satisfying result. By allowing the jury such leeway with respect to demeanor and credibility assessments, the legal system provides for justice—of a sort—to be done and protects the integrity of the law.

124. The law faces a new challenge with the alleged invention of technology for "genetic fingerprinting." It has been touted as a nearly perfect tool for determining whether genetic material found at the scene of a crime (or in some civil context) belongs to the defendant. By taking the guesswork out of identification, it will revolutionize factfinding. Whether it is all that it is cracked up to be, and whether courts can distinguish science fiction from science, remains to be seen. *See* Thompson & Ford, *DNA Typing: Acceptance and Weight of the New Genetic Identification Tests*, 75 VA. L. REV. 45 (1989). But assuming the technique is approximately as flawless as its proponents claim, the optimal gray view of the trial process suggests that the courts will be ambivalent about it. It may be too good. Traditional forensic science, by contrast, may be just what is desired by the law. Its error

3. *Reliability: Actual versus Perceived.* As mentioned earlier in this article, the belief seems widespread that the outcomes of trials are unpredictable. The data about reliability tend to contradict that widespread belief.[125] But the *belief* in uncertainty advances the meta-goal. Lawyers and their clients will settle cases in part because they do not trust their expectations sufficiently to proceed to trial. Yet they have a clear enough idea of the likely outcome that they can and do reach a settlement. At the same time, too much uncertainty makes settlement difficult, creating a pressure for cases to go to trial. Clearly, an optimal balance of pressures to settle and to proceed to trial—of clarity and ambiguity—must be maintained. Generally, parties will settle all but 5 to 10 percent of the cases filed.[126] This rate of settlement may be optimal, or it may need to change in response to changes in society and in the legal system.

At the end of the day, it appears that both the legal process and lawyers are more surefooted than people, including judges and lawyers themselves, realize. And, ironically, the perception of optimal unpredictability contributes to that surefootedness.[127]

---

rates, so long as juries have some sense of them, are just what the courts need: informational help for factfinders that is adequate but not absolute.

125. Law professors commonly tell their students, and lawyers commonly tell their clients, that one can never tell what a judge or (especially) a jury will do. In a draft study of methods of measuring health care quality and sharing that information with consumers (later published as THE QUALITY OF MEDICAL CARE: INFORMATION FOR CONSUMERS, *supra* note 19), authors of one chapter on discipline and malpractice litigation asserted the common wisdom that medical malpractice trials were essentially events whose outcomes were not predictable and meant little. Upon being presented with the sort of data that appear in Table 1 of this article, the authors were both amazed and eager to revise their chapter. Their revisions appear in the completed published study. *See id.* at 134-35.

126. *See* Galanter, *Reading the Landscape of Disputes: What We Know, Don't Know (and Think We Know) About Our Allegedly Contentious and Litigious Society*, 31 UCLA L. REV. 4, 27-28 (1983); Trubek, Sarat, Felstiner, Kritzer & Grossman, *The Costs of Ordinary Litigation*, 31 UCLA L. REV. 72, 86 (1983); Conrad, *The Quantitative Analysis of Justice*, 20 J. LEGAL EDUC. 1, 6-7 (1967).

127. Would the litigation system be able to serve its sampling function if it could decide cases with perfect consistency while counting on the erroneous beliefs of lawyers and the public to keep bringing some, but not most, cases to trial?

*Perhaps.* Even if outcomes were highly predictable and were known to be so by quantitative empirical researchers, those close to the cases might continue to be surprised, if only occasionally, by the verdicts. Those surprises remain more salient than the more numerous non-surprises. (The studies cited at the outset of this article, in Part II.B, address the level of consistency in judge and jury verdicts. These data suggest that outcomes are surprisingly consistent. But clearly there are some surprises.) This situation might be sufficient to insure continuing belief in unpredictability, even of a system that was highly predictable. Moreover, since settlements dominate the system, attorneys have little feedback to learn whether their expectations about the trial outcome were correct or not. It seems that the necessary uncertainty is assured, even if everything were operating at 1.00 reliability and validity.

*But probably not.* Eventually word would filter out that the courts have stopped making mistakes and being unpredictable. Lawyers, perhaps even with the aid of empirical social science studies of the litigation system, would become more and more accurate in their predictions, come to recognize their accuracy, and have confidence in it. After all, physicians cannot know that what they are doing is working by the observation of individual cases; they too must rely on empirical studies with control groups to tell them whether the treatments they are using are working better, worse, or the same as alternative treatments. (In contrast to lawyers—who may disbelieve their and their system's reliability in the face of actual high reliability—doctors seem to believe in the effectiveness of their treatments even when the hard data indicate some of their favorite treatments to be worthless or even iatrogenic. *See generally* COSTS, RISKS, AND BENEFITS OF SURGERY (J. Bunker, B. Barnes & F. Mosteller eds. 1977); THE QUALITY OF MEDICARE CARE: INFORMATION FOR CONSUMERS, *supra* note

4. *Problems with the Theory.* The notion that the law maintains itself in a posture of some optimal range of predictability or unpredictability has two general problems. First, this explanation may not be adequate to explain the law's behavior. Second, other, simpler explanations may be adequate to account for the law's behavior in the situations I have discussed.

For example, one might say that error and foolishness are inevitable among any and all decisionmakers and systems. One need not posit a special goal of instilling optimal error; it will happen. Each such error may have its own idiosyncratic explanation. The search for a systematic, functional explanation for them all might itself be a misdirected search for order in the disorder.

Some unpredictability is inevitably present in life, and especially in the law. So the law need not go to any special trouble to maintain uncertainty. Uncertainty is assured. Moreover, the range of "optimality" would be quite wide. As long as the law's outcomes are not a complete guess or a near certainty, it would seem that the conditions I have suggested are needed will be present.

But assume that I am correct in asserting that factual disputes would dry up and disappear if the law did not keep disputants guessing about factual outcomes. Does it follow in a world of 12(b)(6) motions, declaratory judgments, and demurrers that an inadequately rich picture of the basis of disputes will emerge, that inadequate sampling will take place, and that the loss of refinement in understanding will have a harmful impact on future transactions? Would we have a legal system that was "all normative" and out of touch? We will never know for sure, but one could begin to test the matter by comparing the quality of information about the world that courts get from those cases that are disposed of through pretrial motions and those cases that proceed to trial. Perhaps cases that raise only legal issues present, as an incidental matter, enough of the facts of cases to enable litigation to be an effective sampling system.

## V

## CONCLUSION

This article has focused on procedural and evidentiary devices by which the law seeks to accomplish its factfinding. I have reviewed a broad array of social science findings that inform us about the degree to which the law enhances and restrains accuracy in factfinding, and that suggest means by which the law could improve its trial factfinding capability. The law's trial machinery appears to produce more reliable outcomes than it is widely

---

19.) This contrast may reflect the different needs that the two enterprises have in order to function well. Clients benefit from their lawyers' doubts; patients benefit from their physicians' confidence. Thus, in order for the litigation system as a whole to operate effectively, some real imperfections in both factfinding and application of law must be preserved.

believed to do, yet the law sometimes chooses procedures and evidentiary rules that would seem to yield a poor degree of reliability.

I explore the possibility that these seemingly mistaken choices are a result of the law's need to maintain its level of predictability within some optimal range. If this "optimal gray" explanation has any merit, it has implication for both legal policy and empirical social science research on the legal process. If true, policymakers cannot pursue a strategy of continual improvement in the reliability and accuracy of trial procedures; and empirical researchers need to study the law at a higher level of abstraction (to help the law determine when "accurate" becomes "too accurate") as well as to determine which alternative legal procedures produce more reliable and accurate results.

If the "optimal gray" explanation is incorrect, then the law needs to determine why it sometimes makes policy choices that demonstrably restrain its factfinding capability, raise costs, and endanger other values. At the very least, the law should determine how to make better use of available empirical information to improve its factfinding performance.