

LEGAL REASONING AND SCIENTIFIC REASONING

*Phoebe C. Ellsworth**

I.	DEDUCTIVE AND INDUCTIVE (ANALOGICAL) REASONING	898
II.	BIASES AND HEURISTICS	902
	A. <i>Hindsight Bias</i>	902
	B. <i>Anchoring and Adjustment</i>	903
	C. <i>Confirmation Bias</i>	904
III.	DIFFERENCES BETWEEN LEGAL AND SCIENTIFIC REASONING.....	907
	A. <i>Lack of Opportunity for Empirical Testing</i>	907
	B. <i>The Need for Immediate, Final Decisions</i>	908
	C. <i>Categorical Thinking and Lack of Compromise</i>	913
	D. <i>Individualized Decisions vs. Probabilistic Reasoning, and the Dispositional Bias</i>	915

In my presentation for the 2010 Meador Lectures on Rationality, I chose to compare legal reasoning and scientific reasoning. Both law and science pride themselves on the rationality of their intellectual methods and believe that those methods are designed to analyze questions and reach the correct conclusions by means of reason, free from cognitive or emotional biases. Of course, both law and science often fall short of this ideal at all levels, from the decisions about individual legal cases or scientific studies to the acceptance of general theories. In many ways, the biases that mislead legal and scientific thinkers are similar. But in other ways they are not. Training to think like a lawyer is not quite like training to think like a scientist, and, more important, the circumstances and constraints faced by lawyers and scientists when they undertake the task of solving a problem are quite different.

Another reason for comparing legal and scientific reasoning is a practical one. Scientific evidence has become and is becoming increasingly common in legal cases, and differences in legal and scientific reasoning are a source of frustration to players in both disciplines.¹ Scientific “experts”

* Frank Murphy Distinguished University Professor of Law and Psychology. This paper has been adapted from a lecture delivered as part of the Meador Lectures on Rationality at the University of Alabama School of Law on November 1, 2010.

1. See Susan Haack, *Irreconcilable Differences? The Troubled Marriage of Science and Law*, 72 LAW & CONTEMP. PROBS. 1, 14–21 (2009).

are held in contempt as “hired guns” pushing junk science or as purveyors of impenetrable jargon-ridden testimony that fails to address the question before the court. The experts are frustrated by the seeming barriers to telling “the whole truth” in the context of direct and cross examination. Players on both sides agree that scientific testimony is less helpful than it could be and often blame each other. Fundamental terms have different meanings in the two disciplines: “evidence,” “relevance,” and “reliability.”

The analogy of legal and scientific reasoning was devised and consciously promoted by Christopher Columbus Langdell, Dean of the Harvard Law School from the 1870s until 1895. In his efforts to establish law as a profession rather than a trade, he emphasized the idea that the study of law is a science and that the law faculty, like the faculties of the sciences, is a body of permanent scholars devoted to legal research.

We have also constantly inculcated the idea that the *library* is the proper workshop of professors and students alike; that it is to us all that the laboratories of the university are to the chemists and physicists, the museum of natural history to the zoologists, the botanical garden to the botanists.²

Just as the natural history museum is filled with drawers and jars containing different species to be compared and contrasted with each other and with newly discovered species, the law library would someday be filled with volumes of different kinds of cases to be compared and contrasted with each other and with newly decided cases. Langdell himself compiled the first such volume in 1871, *Selection of Cases in the Law of Contracts*.³

Langdell advocated a type of legal reasoning known as legal formalism. Formalism bore a close resemblance to the biological taxonomic system of Linnaeus, in which classes of plants and animals were organized under phyla, the major divisions of animals or plants (e.g., arthropods); orders under classes (insects); families under orders (e.g., butterflies and moths); genera under families; and species under genera (e.g., Monarch butterfly).⁴ Langdell’s formalism followed the same plan: a pyramid of rules with a very few fundamental “first principles” at the top, such as the Sixth Amendment right to trial by jury, from which mid-level and, finally, a very large number of specific rules could be derived⁵ (e.g.

2. Christopher Columbus Langdell, Address at the Quarter-Millennial Celebration of Harvard University (Nov. 5, 1887), in 3 L.Q. REV. 118, 124 (1887) (emphasis added).

3. M. E. Bathurst, *The Tutorial System: An Experiment in Legal Education*, 9 AM. L. SCH. REV. 695, 695 (1940).

4. Curtis E. A. Karnow, *Rhetoric of Academe*, 41 U.S.F. L. REV. 135, 137 (2006).

5. Thomas C. Grey, *Langdell’s Orthodoxy*, 45 U. PITT. L. REV. 1, 11 (1983).

five-person juries are unconstitutional.)⁶ The legal scientist could move up or down the pyramid. According to Grey, “the fundamental principles of the common law were discerned by induction from cases; rules of law were then derived from principles conceptually; and finally, cases were decided, also conceptually, from rules.”⁷ During the 18th and 19th centuries, scientific and amateur explorers brought back an enormous number of unknown plants and animals, which were closely compared with known species that resembled them, and classified into their proper place in the ruling order. For Langdell, new legal cases were like new species. The method of classifying them by comparing them to existing cases was “the case method,” which revolutionized legal education and is still the legal reasoning taught in first year law courses in order to train students to “think like a lawyer.”⁸

The zoologists and the botanists were a better analogy for Langdell than the physicists and chemists, as taxonomy was still an important part of what they did in the nineteenth century. Chemistry was not mainly concerned with classification but with invention (e.g., dynamite) and new theoretical formulations (e.g., the periodic law). Physicists also focused on new theoretical formulations (e.g., the laws of thermodynamics), but they did not study these by observing real-world phenomena like stoves or treadmills; instead, they created special controlled conditions for studying and measuring the phenomena that they wanted to understand. Even the zoologists and the botanists did much more than merely wait for new specimens to turn up. They cross-fertilized; they travelled to different countries to see how similar species might be modified by different environments; they interfered with the environments in order to see what the effects would be.

Most of what the law does is to decide specific cases. Most of what science does is not. Classifying new species was a bigger part of zoology and botany in Langdell’s time than it is today, but it was not the essence of the science. The essence of the science was to identify underlying processes that led to the differentiation of species. Most social sciences (and if the law is a science at all, surely it must be a social science) are not concerned with specific cases, except when they suggest an exception to a general rule or assumption.

In this regard, the judge is more like the practicing physician or psychiatrist, who is faced with a new patient and must decide on a diagnosis, must make a classification. Psychiatrists use the Diagnostic and

6. See *Ballew v. Georgia*, 435 U.S. 223, 239 (1978).

7. Grey, *supra* note 5, at 19.

8. David D. Garner, *The Continuing Vitality of the Case Method in the Twenty-First Century*, 2000 BYU EDUC. & L.J. 307, 324 (2000).

Statistical Manual of Mental Disorders (DSM) to classify mental disorders.⁹ There are higher-level categories, such as Anxiety Disorders, Depression, Bipolar Disorders, and Schizophrenia.¹⁰ Under Anxiety Disorders there are phobias, panic attacks, post-traumatic stress disorder, generalized anxiety disorder, and others. They function somewhat like the statutes that govern a judge's decisions, and the psychiatrist, like the judge and the 19th century taxonomist, is also informed by the writings of other professionals describing cases they have encountered and how and why they decided them as they did. Like the judge, the psychiatrist is faced with an individual case, a case that is unique but that resembles various other cases that she and others have seen and sometimes described. As judges must justify their decisions in terms of the law, psychiatrists, in communicating their diagnoses to hospitals, insurance companies, and sometimes courts, use the DSM to communicate and justify the patient's diagnosis. As judges' decisions influence their sentencing recommendations, so psychiatrists' diagnoses influence their recommendations about appropriate treatment. Both legal and psychiatric professionals use the case method. In general, I believe that because of this shared interpretive framework the participation of practicing psychiatrists and medical experts in the legal context has been easier than that of scientists who do not deal with individual cases. The use of psychiatric experts is not entirely successful, of course, because the law demands distinctions that are often unimportant or meaningless to the psychiatrists: either/or distinctions, such as sane or insane, intellectually disabled or normal, competent or incompetent. Still, the practice of classifying individual cases according to a set of rules is standard operating procedure in both fields.

I. DEDUCTIVE AND INDUCTIVE (ANALOGICAL) REASONING

Scientific and legal reasoning are typically described as involving two distinct methods: deductive and inductive reasoning. In science, deductive reasoning involves the derivation of specific predictions from a general theory or set of axioms. The general theory leads to more limited hypotheses, and the hypotheses lead to specific predictions which can then be tested by experimentation or observation. For example, a general principle of the theory of evolution is that individuals that are well adapted to their environment succeed by reproducing at greater rates. Based on this general principle, a scientist may hypothesize that when the environment is rapidly changing, evolution will progress at a faster rate. This hypothesis

9. See DIAGNOSTIC AND STATISTICAL MANUAL OF MENTAL DISORDERS (3d ed. 1980).

10. *Id.* at 17–18.

may in turn lead to the prediction that, although black moths far outnumber white moths in the dark smoky environment of a mining town, if a major effort is made to reduce the pollution and clear the air, white moths will come to have an advantage and will soon outnumber the black moths. The implementation of a major cleanup by the town provides an opportunity to test this prediction.¹¹

Law generally begins with the specific case, and the judge looks to the general law (statutes or precedents) from which the solution to this case may be deduced. If a statute prohibits selling alcohol to anyone under the age of twenty-one, and Joe's Liquor Store sells a six-pack of beer to sixteen-year-old Richard, then it follows that Joe's Liquor Store is guilty.

We were all taught about deductive reasoning as "the scientific method" in eighth grade, but in fact, it is fairly uncommon in science and particularly in the social sciences. There are very few widely accepted general theories, and the derivations may be questionable. So, for example, there is a generally accepted principle that people tend to be attracted to others who are similar to them.¹² But in practice there are thousands of possible ways in which people may be similar or dissimilar. Do people like people with the same hair color, size, political beliefs, personality traits, IQ, neuroses, or what? Some derivations from this principle make sense—intellectuals usually like intellectuals more than they like people who never read—but some do not—do mean people like other mean people? Do domineering people like other domineering people?¹³

In law, as in social science, the derivations from widely accepted general principles may not be so clear. The Equal Protection Clause of the Constitution holds that no state may "deprive any person of life, liberty, or property, without due process of law; nor deny to any person within its jurisdiction the equal protection of the laws."¹⁴ History shows that the deductions from this general principle change with time and customs: the Equal Protection Clause did not require that blacks and whites be allowed to sit in the same train car;¹⁵ it prohibited racial segregation in schools;¹⁶ pervasive but "unintentional" race discrimination is not a violation;¹⁷ some

11. See Kenneth R. Miller, *The Peppered Moth—An Update*, EVOLUTION RESOURCES (Aug. 1999), <http://www.millerandlevine.com/km/evol/Moths/moths.html>.

12. See DONN BYRNE, *THE ATTRACTION PARADIGM* 25–31 (1971).

13. See John E. Lydon et al., *Interpersonal Similarity and the Social and Intellectual Dimensions of First Impressions*, 6 SOC. COGNITION 269, 282–85 (1988).

14. U.S. CONST. amend. XIV, § 1.

15. *Plessy v. Ferguson*, 163 U.S. 537, 548 (1896).

16. *Brown v. Bd. of Educ. of Topeka, Shawnee Cnty., Kan.*, 347 U.S. 483, 495 (1954).

17. *Village of Arlington Heights v. Metro. Hous. Dev. Corp.*, 429 U.S. 252, 253 (1977); *McCleskey v. Kemp*, 481 U.S. 279, 351–52 (1987).

forms of affirmative action in education violate the clause, others do not.¹⁸ In law, as in science, the meaning of the general principles changes over time. In science, the general principle may be reworded; in constitutional law, the wording stays the same but the accepted interpretation changes over time. For example, the Supreme Court has explicitly stated that the definition of “cruel and unusual punishment” changes as the values of society develop and mature.¹⁹

In both law and science, the derivation of a particular prediction or decision from the general law is often ambiguous: there may be more than one general principle that is potentially applicable; there may be ambiguities in the general principle; there may be more than one version of the facts or their meaning. In law, attorneys for both sides may propose different deductive arguments—identifying a law and showing how it inevitably leads to an outcome that favors their client in the particular case—but their aim is not to determine the truth but to create a chain of reasoning that best favors their position. The decision maker gets the facts from the attorneys, who emphasize different facts and often different legal precedents. “[T]he law determines what facts are relevant while, at the same time, the facts determine what law is relevant.”²⁰ Thus, even the apparently simplest form of legal reasoning—deciding whether a law covers the facts of the case—is often quite complicated in practice. The sort of deductive reasoning described to schoolchildren is seldom what scientists or judges actually do.

Analogical reasoning involves the comparison of closely related cases to the current case and the placement of the current case in the category of those it most closely resembles, or, occasionally, the creation of a new distinction that differentiates this case from all previous cases. It is “bottom-up” rather than “top-down” reasoning. This is the heart of the case method, introduced by Langdell. Langdell’s original analogy to scientific reasoning was that of a scientist identifying a new species, searching through the drawers of the National History Museum to find the known specimens that most resemble it and classifying it as a member of a known

18. Compare *Grutter v. Bollinger*, 539 U.S. 306, 343 (2003) (holding that the University of Michigan Law School had a compelling interest in promoting class diversity and that its “plus” system did not amount to a quota system that would have been unconstitutional), with *Gratz v. Bollinger*, 539 U.S. 244, 279–80 (2003) (holding the University of Michigan’s “predetermined point allocations” system that awarded 20 points to underrepresented minorities “ensures that the diversity contributions of applicants cannot be individually assessed” and was therefore unconstitutional).

19. *Trop v. Dulles*, 356 U.S. 86, 99–101 (1958) (holding it was unconstitutional for the government to revoke the citizenship of a U.S. citizen as a punishment); see also U.S. CONST. amend. VIII.

20. See STEVEN J. BURTON, AN INTRODUCTION TO LAW AND LEGAL REASONING 141 (2d ed. 1995).

species or a new, closely related species.²¹ Although this still happens in science, it is fairly rare. A more common form of analogical reasoning is to compare a set of experiments that apparently have inconsistent results and identify the variable that distinguishes those with result A from those with result B. For example, sometimes people perform better when other people are around, and sometimes they perform better when they are alone. Numerous studies have provided evidence for both outcomes. In 1965, Robert B. Zajonc explained this paradox by proposing that the presence of others enhances performance on familiar, well-learned tasks (like a running or swimming race) but impairs performance on novel tasks (like finding one's way through a maze).²² The general principle is that arousal facilitates the person's dominant response; when the dominant response is the best one, performance improves; when it is not, performance suffers.²³

In science, this sort of reasoning is reflected in the concepts of convergent and discriminant validity.²⁴ Convergent validity means that many different kinds of tests of the same concept or hypothesis all came out the same way.²⁵ So a researcher who has developed a measure of aggression might show that her measure predicts shouting at people, hurting them, slamming doors, spreading lies about them, and other aggressive behavior. If only "shouting at people" were used as a criterion, a critic could argue that it might be measuring impulsiveness or dominance, not aggression. To demonstrate discriminant validity, the researcher tries to show that the aggression measure is *not* correlated with measures of other states, for example, impulsivity: e.g., it does not correlate with extravagant shopping sprees, risky decisions, or ADHD.²⁶ The term convergent validity is also used when many different research studies using different methods all produce results supporting the same hypothesis. The hypothesis that jurors who favor capital punishment are more likely to find criminal defendants guilty has been tested in surveys, laboratory simulations, and post-trial interviews with actual jurors, using different populations over a period of decades.²⁷ The fact that across all these different methods pro-capital-punishment respondents were more likely than other people to see criminal defendants as guilty greatly increases our confidence that the hypothesis is true, much more so than if all the studies had used the same

21. See Garner, *supra* note 8, at 319; *supra* notes 2–8 and accompanying text.

22. Robert B. Zajonc, *Social Facilitation*, 149 *SCI.* 269, 270 (1965).

23. *Id.*

24. Donald T. Campbell & Donald W. Fiske, *Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix*, 56 *PSYCHOL. BULL.* 81, 81 (1959).

25. *See id.*

26. *See id.*

27. Phoebe C. Ellsworth, *Unpleasant Facts: The Supreme Court's Response to Empirical Research on Capital Punishment*, in *CHALLENGING CAPITAL PUNISHMENT: LEGAL AND SOCIAL SCIENCE APPROACHES* 177, 189–90 (Kenneth C. Haas & James A. Inciardi eds., 1988).

method. Although the terms convergent and discriminant validity are not used in the law, the case method is similar in that the judge or lawyer attempts to define the meaning of a case by looking for factual and legal similarities (convergence) and dissimilarities (discrimination) between this case and related previous cases. This entails defining the universe of possibly applicable cases and deciding which ones match the current case most closely, and which, although apparently similar, do not apply.

II. BIASES AND HEURISTICS

Legal reasoning and scientific reasoning are also subject to errors and biases, often the same errors and biases that affect the reasoning of ordinary people.²⁸

A. Hindsight Bias

Of all the cognitive heuristics and biases that have been studied, hindsight is probably the one most familiar to the public.²⁹ Once people know the outcome of an event, they believe that they would have predicted it in advance. A study by Baruch Fischhoff once asked people to predict the outcome of various current events, such as elections, before they happened, and then asked them after the fact what they had predicted. People tended to say that they had predicted the actual outcome, even when they hadn't.³⁰

In science, this bias may show up as “HARKing”—Hypothesizing After the Results are Known.³¹ Once the scientist sees the results of her work, she comes to believe that they were exactly the results that she predicted and reports them as though they were hypothesized in advance. Sometimes this is done more-or-less strategically, to make it look as though the author had engaged in wise deductive reasoning, but sometimes, the scientists actually believes it was what she had thought all along.³²

Hindsight bias can also occur when the scientist measures an outcome with several different measures and only some of them show the predicted effects.³³ For example, a researcher might test a program designed to

28. See REID HASTIE & ROBYN M. DAWES, RATIONAL CHOICE IN AN UNCERTAIN WORLD: THE PSYCHOLOGY OF JUDGMENT AND DECISION MAKING 18–19 (2001).

29. See Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124, 1124 (1974); see also Scott A. Hawkins & Reid Hastie, *Hindsight: Biased Judgments of Past Events After the Outcomes are Known*, 107 PSYCHOL. BULL. 311, 311 (1990).

30. Baruch Fischhoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXPERIMENTAL PSYCHOL. 288, 295–97 (1975).

31. Norbert L. Kerr, *HARKing: Hypothesizing After the Results are Known*, 2 PERSONALITY & SOC. PSYCHOL. REV. 196, 197 (1998).

32. See *id.* at 204.

33. See *id.*

improve academic motivation in schoolchildren and measure the outcome by looking at hours spent on homework, extra credit assignments, help from the teacher, and academic goals. If some of the measures show the predicted effects and other do not, the researcher may “realize” that the measures that didn’t work were actually inappropriate measures of motivation and so do not disconfirm the hypothesis after all. She may explain away these inconsistent results when she writes up the research, or, more reprehensibly, she may not mention them at all.

The risks of hindsight bias are especially prevalent in legal decision making because, in deciding a case, the judge or jury always knows what happened and what resulted. Their job is to assign blame. The tendency to perceive that the outcome was predictable in advance may lead them to believe that the *defendant* could or should have predicted it. Deciding whether the defendant knowingly caused the damage or whether the outcome was foreseeable in advance is often a crucial element in the choice of a verdict, and hindsight bias suggests that judges and jurors are likely to overestimate the ability to foresee an outcome.

B. Anchoring and Adjustment

When asked to make decisions in ambiguous contexts, people who are given some value, even a completely arbitrary one or one that they know is incorrect, nonetheless tend to latch onto that value and unwittingly use it as an “anchor” in estimating the true value. “[P]eople adjust their estimates from this anchor but nevertheless remain too close to it.”³⁴ For example, if people are asked whether the average annual temperature in New York City is 70°, most people will think that that number is too high. If they are asked whether the average temperature is 40°, most people will think that number is too low. But when asked to guess the actual average annual temperature, people will be influenced by the temperature they were given so that those who heard that the temperature is 70° will estimate that New York is warmer than those who heard that it is 40°. The initial number, even though people reject it, nonetheless serves as an anchor and biases their estimates in that direction.³⁵

Economists often fail to predict major market changes because they are anchored by the status quo; early polls often affect forecasters’ predictions of a candidate’s ultimate success, and so on. In the legal context, a plaintiff’s original demand for damages may seem ridiculously high, but it will nonetheless produce higher damage awards than a more reasonable demand would have. Plausible anchor values influence decision makers

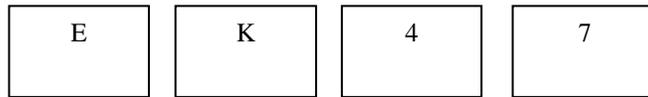
34. HASTIE & DAWES, *supra* note 28, at 99.

35. Tversky & Kahneman, *supra* note 29, at 1128–29.

more than ridiculously extreme values, but even extreme values still exert some anchoring effect. Englich, Mussweiler, and Strack found that judges who were asked whether a sentence of one year was reasonable gave defendants shorter sentences than judges who were asked whether three years was reasonable.³⁶ Jurors may be influenced by a prosecutor's demand for the death penalty to believe that any prison term (e.g., a life sentence) is lenient.

C. Confirmation Bias

Confirmation bias is the tendency to seek, believe, and remember information that agrees with what we already think.³⁷ There are of course motivational reasons for this bias—we want to believe that we perceive the world as it really is and that people who disagree with us are wrong. But the cognitive bias is deeper than a need to preserve our self-esteem and influences our thinking even in situations that are unrelated to our values or our self-image. In a famous experiment, Wason and Johnson-Laird presented participants with four cards, each showing a letter or number.³⁸



Each of the cards has a number on one side and a letter on the other, and the participants were told that “if a card has a vowel on one side, then it has an even number on the other side.”³⁹ Which of the cards would you have to turn over in order to decide whether this is true? The most frequent answer is “E and 4,”⁴⁰ and the next most frequent is “E” alone. Turning over the E can confirm or disconfirm this hypothesis; turning over the 4 provides no information because the statement did *not* say “if a card has an even number then the other side has a vowel.”⁴¹ But turning over the 7 is essential, as the hypothesis would be disconfirmed if there was a vowel on

36. Birte Englich et al., *Playing Dice with Criminal Sentences: The Influence of Irrelevant Anchors on Experts' Judicial Decision Making*, 32 PERSONALITY & SOC. PSYCHOL. BULL. 188, 192–93 (2006).

37. See Raymond S. Nickerson, *Confirmation Bias: A Ubiquitous Phenomenon in Many Guises*, 2 REV. GEN. PSYCHOL. 175, 175 (1998).

38. P. C. WASON & P. N. JOHNSON-LAIRD, *PSYCHOLOGY OF REASONING: STRUCTURE & CONTENT* 173 (1972).

39. *Id.*

40. *Id.*

41. *Id.*

the other side. Few people get this right; apparently testing for disconfirming evidence does not come naturally to people.⁴²

Nickerson has shown that this bias is pervasive in many real-life domains, with people's hypotheses leading them to seek information that is consistent, and ignore or minimize evidence that is inconsistent, with it.⁴³ Rosenthal and Jacobsen told elementary school teachers at the beginning of the school year that some of their students would excel that year.⁴⁴ These "promising" students were actually chosen *at random* by the researchers—there was no evidence that they would do any better than the other students in their classes.⁴⁵ At the end of the year, these students actually did perform especially well.⁴⁶ Apparently, the teachers' expectations translated into their behavior towards the students and affected their evaluations of the students' performances,⁴⁷ in a classic example of a self-fulfilling prophecy.⁴⁸ Men who believe that the woman they are talking to on the phone is attractive infer that she must also be charming and socially skilled and in fact treat her in ways that bring out these qualities in her, compared to men who believe that the woman is unattractive.⁴⁹ Racially prejudiced people's prejudices are confirmed when they interact with minority members, and they treat the minorities in ways that make them perform badly.⁵⁰

In the legal context, Barbara O'Brien gave participants a set of materials resembling a police file in a criminal investigation of a home invasion and shooting, including witness interview reports, ballistic evidence, a photo lineup, a search warrant, and so on.⁵¹ After they had been given the evidence collected in the first half of the file reporting on the early stages of the investigation, half of them were asked to name their prime suspect.⁵² In the materials some evidence pointed to Bill Briggs, who had been fired by the victim several months earlier, and most of the

42. *See id.* at 173–74.

43. Nickerson, *supra* note 37, at 177–78.

44. Robert Rosenthal & Lenore F. Jacobson, *Teacher Expectations for the Disadvantaged*, 218 *SCI. AM.* 19, 19 (1968).

45. *See id.* at 19.

46. *Id.*

47. *See id.* at 22..

48. *See generally* Robert K. Merton, *The Self-Fulfilling Prophecy*, 8 *ANTIOCH REV.* 193 (1948) (discussing whether people's consequent behavior and some of the consequences of that behavior are determined by the ascribed meaning).

49. Mark Snyder et al., *Social Perception and Interpersonal Behavior: On the Self-Fulfilling Nature of Social Stereotypes*, 35 *J. PERSONALITY & SOC. PSYCHOL.* 656, 660–61 (1977).

50. Carl O. Word et al., *The Nonverbal Mediation of Self-Fulfilling Prophecies in Interracial Interaction*, 10 *J. EXPERIMENTAL SOC. PSYCHOL.* 109, 119 (1974).

51. Barbara O'Brien, *Prime Suspect: An Examination of Factors that Aggravate and Counteract Confirmation Bias in Criminal Investigations*, 15 *PSYCHOL. PUB. POL'Y. & L.* 315, 319 (2009).

52. *Id.* at 320.

participants named him.⁵³ The other half were not asked to name anyone.⁵⁴ Then all participants read the second half of the file which included some additional evidence consistent with Briggs's guilt, but also some evidence that cast doubt on it, and some evidence that pointed to other possible suspects.⁵⁵ Compared to participants who were not asked to name anyone, participants who had named Briggs tended to remember more facts that were consistent with Briggs's guilt, suggested new lines of investigation focused on Briggs rather than on other possibilities, interpreted ambiguous evidence as consistent with Briggs's guilt, requested more reports focused on Briggs, and evaluated inconsistent evidence in a way that confirmed their suspicions.⁵⁶ O'Brien concluded that confirmation bias, in this case confirmation of the guilt of an initial suspect, may be an important contribution to false convictions.⁵⁷ In the legal context, this sort of bias has been referred to as "tunnel vision."⁵⁸

Scientists are certainly susceptible to confirmation bias. Many empirical scientists in all fields design studies to test a single hypothesis, which can lead them to focus only on the results that support that hypothesis, or if none of the results support the hypothesis, to decide that there was a problem with the method they used, not with the hypothesis. Scientists and philosophers of science have, over the centuries, argued that the right strategy is to seek methods or contexts that could disconfirm the hypothesis⁵⁹ or, better yet, to test multiple competing hypotheses simultaneously, but our bias towards hypothesis confirmation makes it difficult to put these recommendations into practice.⁶⁰ Like the participants in O'Brien's research, the scientist attends to findings that confirm her hypothesis, remembers them, explains away findings that do not, and designs the next study so as to provide further confirmation. Doctors and clinical psychologists are affected by their preconceptions and stereotypes in diagnosing patients. O'Brien discusses the relation of her work to police and prosecutors,⁶¹ but other actors in the legal system, such as judges and

53. *Id.* at 319–20.

54. *Id.* at 320.

55. *Id.*

56. *Id.* at 328.

57. *Id.* at 331.

58. Keith A. Findley & Michael S. Scott, *The Multiple Dimensions of Tunnel Vision in Criminal Cases*, 2006 WIS. L. REV. 291, 292 (2006).

59. See generally KARL POPPER, *THE LOGIC OF SCIENTIFIC DISCOVERY* (Routledge 2002) (1935).

60. See T. C. Chamberlin, *The Method of Multiple Working Hypotheses*, 15 SCI. 92 (1890), reprinted in 148 SCI. 754, 754–55 (1965); see also Phoebe C. Ellsworth & Richard Gonzales, *Questions and Comparisons: Methods of Research in Social Psychology*, in THE SAGE HANDBOOK OF SOCIAL PSYCHOLOGY 24 (Michael A. Hogg & Joel Cooper eds., 2003).

61. O'Brien, *supra* note 51, at 318.

jurors, are susceptible as well. Confirmation bias is a pervasive human tendency.

In the courtroom, the use of the adversary system may provide some protection against confirmation bias. The judge or the jury almost always has to consider two competing hypotheses, one presented by the plaintiff or prosecutor and the other by the defendant. The scientist often has a single hypothesis and isn't *forced* to consider any alternatives, despite the recommendation of specialists in the scientific method.⁶² However, the adversary system is by no means a complete protection against confirmation bias. As Dan Simon has argued, the judge's initial uncertainty eventually fades as he realizes how he will probably decide the case, and the arguments that seemed to favor the other side dwindle into insignificance due to pressures toward consistency and coherence.⁶³ The result is that most judicial outcomes are written as though there was never any uncertainty; the decision was inevitable and the arguments for it unequivocal. Jurors may be less confident in their decisions because they often have to deal with arguments for different verdicts throughout the decision-making process, and even when unanimity is achieved, not all jurors may be equally confident of the correctness of the final verdict.

III. DIFFERENCES BETWEEN LEGAL AND SCIENTIFIC REASONING

Despite their similarities, scientific reasoning and legal reasoning differ in fundamental ways. They are bound by different rules, subject to different constraints, and driven by different goals. These differences are not simply a matter for abstract intellectual analysis; as scientific evidence has become increasingly integral to legal decision making, the "deep tensions between the goals and values of the scientific enterprise and the culture of the law," as Susan Haack put it, raise obstacles on the path towards truth and arouse frustration in both experts and the legal decision makers.⁶⁴

A. Lack of Opportunity for Empirical Testing

Perhaps the most important difference between scientific and legal problem solving is that when the evidence is ambiguous, the natural response of the scientist is to collect new data to illuminate the question. Scientists are not constrained by the evidence at hand but can design and carry out new observations or experiments to produce new evidence to

62. Campbell & Fiske, *supra* note 24, at 104.

63. Dan Simon, *A Psychological Model of Judicial Decision Making*, 30 RUTGERS L.J. 1, 19–21 (1998); *see also* Duncan Kennedy, *Freedom and Constraint in Adjudication: A Critical Phenomenology*, 36 J. LEGAL EDUC. 518 (1986).

64. Haack, *supra* note 1, at 2.

address the ambiguities. This avenue is closed to judges. Judges must work with the information presented to them in court, or in briefs, or both; in the opinions of previous judges who have decided similar questions; in the law itself, sometimes scholarly writings; and in, of course, their own common-sense knowledge of the world. They can select from these sources and interpret them in new ways, but they cannot create information that is not already in the system.

Attorneys can create new information. Although the usual method is to ask experts to testify about procedures they have already carried out or studies they have already conducted, occasionally attorneys will work with experts to conduct new research. For example, the expert may be asked to conduct a survey in a community to find out whether the jury pool is representative or whether in a highly publicized case a significant proportion of the citizens in the community has formed strong opinions that might impair their ability to judge the case fairly. Psychological or linguistic experts may conduct research to determine whether a particular set of judicial instructions is comprehensible to potential jurors⁶⁵ or whether a particular photo lineup is suggestive.⁶⁶

Although the quality of this kind of research can be as high as that of anything in the field, it is important to remember that the attorney's goal is not to discover the truth but to win the case. If the results of the study he has requested, however valid, do not contribute to the goal of winning, they will not appear in court. If the expert objects that scientific truth demands that they be presented, then the expert will not appear in court either. There are always other experts, and the attorney has no obligation to choose the best scientist or the most valid testimony. In fact, attorneys may argue that they are ethically bound to use the expert whose testimony will most help their client, regardless of the quality of the science.

B. The Need for Immediate, Final Decisions

Scientists have a great luxury and privilege that is not available to judges: when the evidence is incomplete or ambiguous, they can say, "I don't know." Scientists can create new information, and before they have managed to do so, they can say, "No answer is possible at the present time. We must wait for new information." Neither option is open to judges. They cannot conduct new research to bring new evidence to bear on the case, and they cannot say that there is not enough evidence to make a decision. Even

65. See Shari Seidman Diamond & Judith N. Levi, *Improving Decisions on Death by Revising and Testing Jury Instructions*, 79 JUDICATURE 224, 230 (1996).

66. See Gary L. Wells et al., *Accuracy, Confidence, and Juror Perceptions in Eyewitness Identification*, 64 J. APPLIED PSYCHOL. 440, 440-43 (1979); see also ELIZABETH F. LOFTUS, EYEWITNESS TESTIMONY 150-51 (1979).

when they feel that there is not enough evidence to make a confident decision, they must make a decision anyway. In the process of making that decision, their confidence increases, and they often forget that they were ever uncertain.⁶⁷ A scientist who ponders a problem and concludes that no answer is possible at the present time may be admired; a judge who reached such a conclusion would be considered incompetent: “Because the legal system aspires to resolve disputes promptly, the scientific questions to which it seeks answers will often be those for which all the evidence is not yet in.”⁶⁸

The modern era of empirical research in psychology and law was in part stimulated by a Supreme Court decision that many in the social science community considered premature.⁶⁹ In 1970, in the case of *Williams v. Florida*, the Court held that six-person juries, which were rare at the time, were constitutionally permissible, citing “experiments” in support of its claim that “the reliability of the jury as a factfinder hardly seems likely to be a function of its size.”⁷⁰ It also claimed that jury representativeness would not be compromised by the six-person jury because “in practice the difference between the [twelve]-man and the six-man jury in terms of the cross-section of the community represented seems likely to be negligible.”⁷¹ In 1972, this time citing a scant amount of research, the Court decided that the requirement that juries reach unanimous verdicts was not constitutionally required, as non-unanimous juries would deliberate as thoroughly and would achieve the same level of certainty beyond a reasonable doubt as juries that were required to reach unanimous verdicts.⁷²

Members of the social science community were surprised and dismayed that the Court could allow such significant changes in the jury system on the basis of such flimsy evidence about human behavior.⁷³ These decisions stimulated an explosion of research on juries.⁷⁴ However, the *Williams* and *Johnson* decisions set precedents, and the new research, showing that both smaller size and lack of a unanimity requirement did in fact diminish the quality of jury decision making, appeared to be an

67. See Simon, *supra* note 63, at 92–94.

68. Haack, *supra* note 1, at 16 (emphasis and internal citation omitted).

69. Phoebe C. Ellsworth & Robert Mauro, *Psychology and Law*, in 2 THE HANDBOOK OF SOCIAL PSYCHOLOGY 684, 685 (Daniel T. Gilbert et al. eds., 4th ed. 1998).

70. *Williams v. Florida*, 399 U.S. 78, 100–01 (1970).

71. *Id.* at 102.

72. *Johnson v. Louisiana*, 406 U.S. 356, 360–63 (1972); *Apodaca v. Oregon*, 406 U.S. 404, 410–12 (1972).

73. See Richard O. Lempert, *Uncovering “Nondiscernable” Differences: Empirical Research and the Jury-Size Cases*, 73 MICH. L. REV. 643, 644–45 (1975); see also Hans Zeisel, . . . *And Then There Were None: The Diminution of the Federal Jury*, 38 U. CHI. L. REV. 710, 712–15 (1971).

74. See REID HASTIE ET AL., *INSIDE THE JURY* (The Lawbook Exch., Ltd. 2002) (1983); see also VALERIE P. HANS & NEIL VIDMAR, *JUDGING THE JURY* (1986).

example of closing the barn door after the horse has already escaped. The research was accurately discussed by Justice Blackmun in *Ballew v. Georgia* (1978), when the Court held that five-person juries were unconstitutional because they compromised the representativeness of the jury, the quality of the deliberation, and the accuracy of the verdict.⁷⁵ While the social scientists may have been pleased to be so extensively cited, the actual outcome of the case was not supported by the research record. Almost all of the research compared six-person and twelve-person juries and concluded that twelve-person juries were superior.⁷⁶ No research actually addressed the question of differences between five-person and six-person juries,⁷⁷ and it is doubtful that any significant differences would be found. Nonetheless, following *Ballew*, six-person juries are constitutional, and five-person juries are not.⁷⁸ From a scientific point of view, the obvious conclusion would be that *Williams* should be overturned, but the force of precedent, and the high value the law sets on finality, made that an impossible outcome at the time.

Once in a while the Supreme Court has handed down a decision that has acknowledged that the current evidence was insufficient to support a constitutional claim but suggested that further research might provide more decisive evidence that might be convincing. In *Witherspoon v. Illinois* (1968), the appellant claimed that excluding all opponents of the death penalty from jury service in capital cases resulted in harsher, more conviction-prone juries, who would be biased toward guilty verdicts, compared to juries drawn from a representative cross-section of the population.⁷⁹ The empirical record was thin, and the Court rightly decided that it was “too tentative and fragmentary” to provide convincing evidence of an unconstitutional bias in the procedure.⁸⁰ However, in an uncharacteristically scientific move, the majority appeared to acknowledge that the question was an empirical one and that the decision could be revised if additional research supported *Witherspoon*’s claim.⁸¹ That this reasoning was unusual for the Court is illustrated by the fact that when abundant new supportive research was presented to the Court in 1986 in *Lockhart v. McCree*, there, the Court (now with a very different

75. 435 U.S. 223, 231–39, 245 (1978).

76. *Id.* at 242–43.

77. *Id.* at 234.

78. *Id.* at 239, 242.

79. 391 U.S. 510, 515–17 (1968).

80. *See id.* at 517–18.

81. *Id.* at 518 (stating that “[i]n light of the presently available information, we are not prepared to announce a *per se* constitutional rule”); *see also id.* at 520 n.18.

composition) held that the practice was constitutional and stated that the issue was not after all an empirical one.⁸²

The law's need for immediate decisions also has consequences for the quality of scientific evidence presented in the courtroom. When the empirical evidence in support of an argument is weak, many of the most qualified experts will refuse to testify or will insist upon providing truthful testimony about the unreliability of the evidence, in which case the attorney, having the best interests of the client in mind, will not use them. However, not all "experts" are so scrupulous, and the attorney can often find someone who is willing to testify to the truth of a proposition that is generally rejected by the scientific community, scientists

from the farther ends of the spectrum of scientific opinion: those ready to give a confident answer before others think any answer is warranted, those more scientifically radical, or more scientifically conservative, than most of their colleagues, those whose views have become dogmatically entrenched in the course of their involvement in the litigation process—and, no doubt, a few outright cranks and a few outright whores.⁸³

It might be argued that the adversarial process provides protection against these disreputable experts: the other side can call in an expert to refute the unscientific testimony. The Court in *Frye v. United States*⁸⁴ held that research evidence must be "generally accepted" by the relevant scientific community in order to be admissible in court, and *Daubert v. Merrell Dow Pharmaceuticals*⁸⁵ held that acceptance is an important factor for admissibility. But judges rarely have the education or knowledge to assess the general level of acceptance of scientific information. The plaintiff's expert will say that it is generally accepted; the defendant's will say that it is not. And general acceptance can be redefined by redefining the relevant field. Small specialized fields of practitioners outside of the scientific community—such as handwriting analysts, advocates for some forms of alternative medicine, and bite-mark analysts, to name a few—may recognize "experts" in their field, but the whole field may be regarded as suspect by the larger scientific community. For years courts relied on "arson experts" to analyze the physical evidence in buildings that had been burned down and provide expert testimony as to whether or not the fire was the result of arson. Accumulated lore about patterns of cracked glass and movements of the fire consistent with "accelerants" such as kerosene was

82. See 476 U.S. 162, 165, 172 (1986).

83. Susan Haack, *Of Truth, in Science and in Law*, 73 BROOK. L. REV. 985, 1001–02 (2008).

84. 293 F. 1013, 1014 (D.C. Cir. 1923).

85. 509 U.S. 579, 596 (1993).

widely accepted among this community of experts, but rigorous comparisons of the residues of natural and arson-caused fires showed that none of the supposedly diagnostic criteria actually distinguished between them.⁸⁶

Finally, in cases where the generally accepted scientific answer is simply that “we don’t know—there is not enough evidence,” often neither side will be interested in calling the expert because statements of uncertainty are rarely likely to be persuasive. In the case of *Barefoot v. Estelle*, even the Supreme Court was not persuaded by strong scientific consensus that a question was unanswerable given the present state of knowledge.⁸⁷ In some states, one of the factors that jurors are supposed to consider in deciding between the death penalty and life imprisonment is whether the defendant is likely to be a danger to society in the future, to continue to commit violent crimes.⁸⁸ If their answer to this question is yes, the death penalty is warranted. Some expert witnesses, such as the notorious Dr. Grigson in Texas, could be counted on to testify in every case that they had assessed the defendant and were perfectly confident that he would definitely continue a life of violence no matter what the circumstances.⁸⁹ In *Barefoot*, the American Psychiatric Association filed an amicus brief detailing the strong consensus among doctors and social scientists that it was impossible to predict future dangerousness with any accuracy and that it should not be permissible as an aggravating factor because any evidence presented in court would be meaningless.⁹⁰ In fact, the data indicated that predictions of future dangerousness were usually wrong.⁹¹

Nonetheless, the Supreme Court held that it was constitutional for juries to consider future dangerousness in deciding between life and death and for experts to make predictions about the likelihood that the defendant would pose a continuing threat to society.⁹² The majority dismissed the psychiatric evidence that such testimony was useless, saying that “[n]either petitioner nor the Association suggests that psychiatrists are *always* wrong with respect to future dangerousness, only most of the time.”⁹³ The Justices

86. See TEXAS FORENSIC SCIENCE COMMISSION, REPORT ON THE WILLINGHAM/WILLIS INVESTIGATION 18–28 (2011), available at <http://www.fsc.state.tx.us/documents/FINAL.pdf>; see generally DAVID J. ICOVE & JOHN D. DEHANN, FORENSIC FIRE SCENE RECONSTRUCTION (2d ed. 2008) (describing the need for the scientific method in fire scene investigations).

87. See 463 U.S. 880, 897–900 (1983).

88. See *id.* at 896.

89. See *id.* at 905 n.11, 936–37; see also *supra* Haack, note 1, at 17.

90. See *Barefoot*, 463 U.S. at 920.

91. See *id.* (“The APA’s best estimate is that *two out of three* predictions of long-term future violence made by psychiatrists are wrong.”).

92. *Id.* at 905–06.

93. *Id.* at 900 (emphasis added).

seemed to reason that predicting the future violence of defendants is sufficiently important for capital sentencing; if a legislature stipulates it, the jury must be allowed to do it, even when the strong consensus among the experts is that it cannot be done.

C. Categorical Thinking and Lack of Compromise

The need to reach final decisions in individual cases also encourages categorical thinking: a defendant is either liable or not liable, sane or insane, a danger to society or not. Scientists, especially social scientists, are more likely to think in terms of continuous variables; there is always a grey area between the sane and the insane, the dangerous and the safe, and the deliberate or unintentional behavior. In dealing with people in these grey areas, the task is to assess the individual and the circumstantial pressures and to come up with an individually nuanced explanation, and if one is a psychiatrist or a clinical psychologist or some other kind of counselor, an individualized plan of treatment. But a judge has to make a decision, and this has to be a categorical decision: liable or not liable; guilty or not guilty; guilty of first degree murder, second degree murder, manslaughter, or not guilty; and so on. A judge can sometimes temper this stark decision by leniency in setting damages or deciding sentences, but there are few formal opportunities for compromise decisions.⁹⁴

This general requirement of “either–or” decisions in deciding the ultimate outcome of a case seeps into other evidentiary decisions, including the presentation and interpretation of expert evidence. Experts in the field may agree about most of the data, with disputes mainly about future implications or secondary findings, but this comforting scientific consensus is not likely to be apparent in court. The pressure towards categorical thinking in the context of an adversarial system means that experts who take unambiguous, often extreme, positions are favored, and even experts who agree with other experts on most issues but differ at the edges are encouraged to ignore the vast areas of agreement and to emphasize the differences.⁹⁵ An expert who wants to present an overview of the whole question, emphasizing both areas of agreement and areas of disagreement, may be stymied, because the attorney will only ask questions about the disagreements. This is one reason that junk scientists are favored by attorneys more often than they should be, and serious scientists are frustrated and disillusioned by their attempts to provide useful evidence in court.

94. The principle examples are plea bargains and pretrial settlements.

95. See Haack, *supra* note 1, at 17–18.

Daubert does not provide much protection against this problem because judges deciding admissibility under *Daubert* also make yes-or-no decisions about the validity of each expert's testimony. An expert who agrees with an opposing expert except at the edges is likely to be seen as having little to add. An expert whose research fails to address crucial aspects of the question may be ruled inadmissible for that reason: her testimony does not point to an unequivocal yes-or-no answer. Or a judge may reject an expert because of some imperfection in the research. But in science, certainty is often achieved by examining a coalescence of many studies, using many methods, each with different flaws and different strengths, so that one's methodological strengths compensate for the methodological weaknesses of others, ultimately producing a "convergent validity" that is stronger than the validity of any single study.⁹⁶ By making a categorical "either-or" decision on each expert witness's contribution, a strong consensus among experts about the ultimate conclusion may be overlooked.⁹⁷

This piecemeal, categorical decision making is also evident at the appellate level. In the case of *Lockhart v. McCree*, the question of whether juries made up exclusively of people who would be willing to sentence a person to death would be biased against defendants in deciding *guilt* came before the Supreme Court.⁹⁸ The Court had rejected this claim in 1968 in *Witherspoon* because the empirical evidence for a bias toward guilt was too tentative and fragmentary.⁹⁹ By 1986 when *Lockhart* was decided, the evidence consisted of fifteen empirical studies, conducted over a thirty-year period, using samples from different geographic areas and different demographic groups, and multiple methods: surveys, experiments, simulations, and observations of actual jurors. All of the research converged on the conclusion that death-qualified jurors were more favorable to the prosecution and were more likely to vote guilty than the citizens who were excluded, and jury deliberation did not erase the bias.¹⁰⁰

Justice Rehnquist, writing for the majority, examined the fifteen studies one by one, finding a flaw in each and discarding it from the set, until only one was left,¹⁰¹ and "[s]urely" he concluded, a constitutional decision "should not be based on . . . [a] lone study . . ."¹⁰² Although the American Psychological Association's amicus brief, the lower court cases, and Justice

96. See Campbell & Fiske, *supra* note 24, at 81.

97. Haack, *supra* note 83, at 1003-04.

98. See 476 U.S. 162, 184-85 (1986).

99. *Witherspoon v. Illinois*, 391 U.S. 510, 517-18 (1968).

100. See Ellsworth, *supra* note 27, at 189-91.

101. *Lockhart*, 476 U.S. at 168-72.

102. *Id.* at 172-73.

Marshall's dissent clearly described the concept of convergent validity,¹⁰³ the Court majority managed to overlook it, instead making a categorical "valid or invalid" decision about each study on different grounds.

*D. Individualized Decisions vs. Probabilistic Reasoning,
and the Dispositional Bias*

In the social sciences, reasoning and evidence are usually probabilistic. For example, researchers may examine hundreds or thousands of cases to discover whether racial or gender discrimination exists in a company,¹⁰⁴ or an organization, or a procedure such as jury selection,¹⁰⁵ or capital sentencing¹⁰⁶ to find that women or African Americans or some other group are hired less often, chosen for the jury venire less often, or sentenced to death more often than white men. This is not the question a judge is trying to answer, though the evidence may be relevant. The judge wants to know whether this particular woman or African American was discriminated against because of gender or race in this particular case. This is not a question scientists know how to answer.¹⁰⁷ As mentioned earlier, this is one reason that psychiatrists, like other doctors, have had an easier experience as expert witnesses than research social scientists or epidemiologists. Their question—"what is the nature of this person?"—is the same as the law's question, and one of the most pervasive (and least visible) uses of social science expertise in the legal system is the evaluation of the "fitness, dangerousness, sanity, truthfulness, or emotional well-being of individuals."¹⁰⁸ Aggregate and probabilistic data showing a systematic pattern of racial discrimination, or a risk factor for disease (smoking), or the lack of one (Bendectin) are suspect because they cannot tell us whether race, or smoking, or Bendectin was definitely the cause of this particular person's misfortune.¹⁰⁹

The courts have become more receptive to aggregate statistical data over the past half-century. Data of this kind figure regularly in employment

103. See *id.* at 186–88.

104. See Word et al., *supra* note 50, at 119; Susan T. Fiske, *Stereotyping, Prejudice, and Discrimination*, in 2 THE HANDBOOK OF SOCIAL PSYCHOLOGY 357, 357–58 (Daniel T. Gilbert et al. eds., 4th ed. 1998).

105. See generally *Castenada v. Partida*, 430 U.S. 482, 495–96 (1977) (holding that *Castenada* made out a prima facie case of purposeful discrimination by presenting census statistics that showed the grand jury selection list had a disproportionately low number of Mexican-Americans).

106. See DAVID C. BALDUS ET AL., EQUAL JUSTICE AND THE DEATH PENALTY: A LEGAL AND EMPIRICAL ANALYSIS 312–14 (1990); see also Ellsworth & Mauro, *supra* note 69, at 719.

107. Haack, *supra* note 83, at 985–87.

108. Phoebe C. Ellsworth & Julius Getman, *Social Science in Legal Decision Making*, in LAW AND THE SOCIAL SCIENCES 581, 588 (Leon Lipson & Stanton Wheeler eds., 1986).

109. See *Daubert v. Merrell Dow Pharmaceuticals, Inc.*, 509 U.S. 579, 582–83 (1993) (discussing petitioner's claims that Bendectin caused birth defects).

discrimination cases, for example, but there is still resistance. When a criminal case is based on an eyewitness identification, some judges allow an expert to testify about the factors that impair the accuracy of an eyewitness identification, but others do not, often rejecting the expert on the grounds that he or she did not examine *this particular* eyewitness but can only talk about the factors that affect most eyewitness most of the time.

In the case of *McCleskey v. Kemp*, petitioners presented a study by David Baldus and his colleagues in which they analyzed data on more than 400 variables in over 1,000 Georgia homicide cases and found that race was the only variable that could explain the discrimination against defendants who killed white victims.¹¹⁰ These defendants were more likely to be sentenced to death than those who killed black victims, and blacks who killed whites were the most likely to be sentenced to death.¹¹¹ In a 5–4 decision, the Court held that this sort of aggregate statistical evidence is inherently “insufficient to support an inference that any of the [individual] decisionmakers in McCleskey’s case acted with discriminatory purpose.”¹¹² The only evidence that would be sufficient to show that McCleskey’s death sentence violated the Constitution would be evidence that proved either that the particular jurors in his case were motivated by racial bias or that the Georgia legislature was motivated by racial bias when it enacted the death penalty legislation.¹¹³ This standard created an essentially insuperable barrier to proof of racial discrimination in capital sentencing. Rachel Moran argues that this emphasis on disparate intent (for which aggregate social science data are irrelevant) instead of disparate impact (for which such data are important) has actually increased under the Rehnquist and Roberts Courts.¹¹⁴

I believe that the legal system’s distrust of statistics is not just a matter of lack of training in math or empirical methods. A statistically significant finding implies that other people, perhaps *most* people, would have behaved the same way in the same circumstances, suggesting that the circumstances largely determined the behavior. Much of the law is grounded in a belief in personal responsibility and free will, whereas the social scientists work from a much more deterministic set of assumptions, documenting cultural, sociological, psychological, and especially in psychology, situational influences on behavior.¹¹⁵

110. See 481 U.S. 279, 286–87 (1987).

111. *Id.* at 287.

112. *Id.* at 297.

113. See *id.*

114. Rachel F. Moran, *What Counts As Knowledge? A Reflection on Race, Social Science, and the Law*, 44 LAW SOC’Y REV. 515, 541–43 (2010).

115. See LEE ROSS & RICHARD E. NISBETT, *THE PERSON AND THE SITUATION: PERSPECTIVES OF SOCIAL PSYCHOLOGY* 187–96 (1991).

Of course, the law recognizes that personal responsibility can be a matter of degree and has developed an elaborate set of definitions of different levels of responsibility, including malice aforethought and deliberation, intention, knowledge, recklessness, and negligence. But it ignores, and perhaps *must* ignore, the social psychological evidence that the situation plays a far greater role in determining people's behavior than their personal preferences or dispositions. In fact, situational pressures shape people's preferences and intentions.¹¹⁶ Not only legal actors, but people in general, are biased towards internal dispositional explanations for other people's behavior (but not their own),¹¹⁷ a bias so pervasive that it is known as "the fundamental attribution error."¹¹⁸

The social science literature is full of studies demonstrating the dominance of situational forces over personal dispositions and people's tenacious belief in personal dispositions. In one famous example, Darley and Batson studied seminary students who were assigned to give a talk in a nearby building.¹¹⁹ They first filled out a measure of the strength and personal meaning of their religious beliefs. Half of the seminarians were assigned to talk about the parable of the Good Samaritan, the other half about seminary jobs.¹²⁰ In addition, some were told that they were already late for the talk, and some were given more time.¹²¹

On the way to the other building, each seminarian passed a man who was slumped in a doorway, moaning and coughing.¹²² The question was: which ones would stop and help him? Those with the strongest religious beliefs? Those who were thinking about the Good Samaritan? Or those who had plenty of time? Most people predicted that the most religious students and the ones who were thinking about the Good Samaritan would be the most likely to stop and help the man.¹²³ But in fact, these dispositional factors had little or no effect. What mattered was whether they were in a hurry: sixty-three percent helped when they had the extra time, but only ten percent helped when they were in a hurry.¹²⁴

When the law *does* take account of situational forces, as in cases of necessity and duress, they are usually so extreme as to be the stuff of melodrama: someone asks the bank teller to hand over the money because

116. *Id.* at 187.

117. *Id.* at 189.

118. *See id.* at 130–32.

119. John M. Darley & C. Daniel Batson, "From Jerusalem to Jericho": A Study of Situational and Dispositional Variables in Helping Behavior, 27 J. PERSONALITY & SOC. PSYCHOL. 100, 102 (1973).

120. *See id.* 102–03.

121. *Id.* at 103–04.

122. *Id.* at 104.

123. *See id.* at 102.

124. *See id.* at 105.

another person is holding a gun to her head or breaks into a cabin in the wilderness because she is freezing to death. The extremity of these examples suggests the law's failure to recognize the ubiquitous power of the situation in all aspects of people's daily lives.¹²⁵ Even when the law does make exceptions to attributions of responsibility, these exceptions usually have to do with aspects of the person (insanity or youth), rather than aspects of the situation.

The judge's task is to assign blame, and the assumption of free will is necessary to carrying out the judicial function. But the fact that it is necessary does not mean that it accurately reflects the blameworthiness of the individual, and perhaps some awareness of the fundamental attribution error might temper the punitive nature of American sanctions.

125. See Jon Hanson & David Yosifon, *The Situation: An Introduction to the Situational Character, Critical Realism, Power Economics, and Deep Capture*, 152 U. PA. L. REV. 129, 153 (2003); Lee Ross & Donna Shestowsky, *Contemporary Psychology's Challenges to Legal Theory and Practice*, 97 NW. U. L. REV. 1081, 1087-88 (2003).