Cornell Law Review

Volume 93
Issue 1 *November 2007*

Article 9

Blinking on the Bench: How Judges Decide Cases

Chris Guthrie

Jeffrey J. Rachlinski

Andrew J. Wistrich

BLINKING ON THE BENCH: HOW JUDGES DECIDE CASES

Chris Guthrie,† *Jeffrey J. Rachlinski*†† *& Andrew J. Wistrich*†††

How do judges judge? Do they apply law to facts in a mechanical and deliberative way, as the formalists suggest they do, or do they rely on hunches and gut feelings, as the realists maintain? Debate has raged for decades, but researchers have offered little hard evidence in support of either model. Relying on empirical studies of judicial reasoning and decision making, we propose an entirely new model of judging that provides a more accurate explanation of judicial behavior. Our model accounts for the tendency of the human brain to make automatic, snap judgments, which are surprisingly accurate, but which can also lead to erroneous decisions.[1] *Equipped with a better understanding of judging, we then propose several reforms that should lead to more just and accurate outcomes.*

† Associate Dean for Academic Affairs and Professor of Law, Vanderbilt Law School.
†† Professor of Law, Cornell Law School.
††† Magistrate Judge of the United States District Court for the Central District of California. The authors appreciate feedback from faculty participating in workshops at Cornell Law School, the University of California-Berkeley Law School, the University of Illinois College of Law, and the University of Iowa College of Law. For comments on earlier drafts, the authors thank Susan Bandes, Tracey George, John Goldberg, and Dan Kahan.

1 *See generally* MALCOLM GLADWELL, BLINK: THE POWER OF THINKING WITHOUT THINKING (2005) (discussing the role of automatic decisions made "without thinking").

> "I fear the judge who is too sure of himself, who reaches his decision quickly, jumping immediately to conclusions without deliberation or repentance."[2]

INTRODUCTION

How *do* judges judge? The answer to this seemingly simple question has proved surprisingly elusive.

The two venerable models of judging—the formalist and realist models—offer contrasting responses. According to the formalists, judges apply the governing law to the facts of a case in a logical, mechanical, and deliberative way.[3] For the formalists, the judicial system is a "giant syllogism machine," and the judge acts like a "highly skilled mechanic."[4] Legal realism, on the other hand, represents a sharp contrast. According to the realists, judges follow an intuitive process to reach conclusions which they only later rationalize with deliberative reasoning.[5] For the realists, the judge "decides by feeling, and not by judgment; by 'hunching' and not by ratiocination" and later uses deliberative faculties "not only to justify that intuition to himself, but to make it pass muster."[6] Neither model has proved satisfactory. Judges surely rely on intuition, rendering a purely formalist

---

2 PIERO CALAMANDREI, EULOGY OF JUDGES 21 (John Clarke Adams & C. Abbott Phillips, Jr. trans., 1942).

3 *See* Burt Neuborne, *Of Sausage Factories and Syllogism Machines: Formalism, Realism, and Exclusionary Selection Techniques*, 67 N.Y.U. L. REV. 419, 421 (1992).

4 *Id.* For other depictions of the formalist approach to judging, see, for example, BRIAN BIX, JURISPRUDENCE: THEORY AND CONTEXT 183 (Sweet & Maxwell Ltd. 4th ed. 2006) (1996) (describing formalist judicial decision making as "a nearly mechanical, nearly syllogistic move from basic premises to undeniable conclusion"); Brian Leiter, *Positivism, Formalism, Realism*, 99 COLUM. L. REV. 1138, 1145–46 (1999) (book review) (describing legal formalism as a "descriptive theory of adjudication according to which (1) the law is rationally determinate, and (2) judging is mechanical. It follows, moreover, from (1), that (3) legal reasoning is *autonomous*, since the class of *legal* reasons suffices to justify a unique outcome; no recourse to non-legal reasons is demanded or required.").

5 *See* Joseph C. Hutcheson, Jr., *The Judgment Intuitive: The Function of the "Hunch" in Judicial Decision*, 14 CORNELL L.Q. 274, 285 (1929).

6 *Id.*; *see also* JEROME FRANK, LAW AND THE MODERN MIND 111–12 (Anchor Books, Peter Smith 1970) (1930) (characterizing Hutcheson's depiction of judging "as an approximately correct description of how all judges do their thinking"); Hutcheson, *supra* note 5, at 278–79 (describing his decision-making process as a judge, Hutcheson writes, "I . . . give my imagination play, and brooding over the cause, wait for the feeling, the hunch—that intuitive flash of understanding . . . ."). Note that scholars often distinguish between the Frank/Hutcheson approach to realism and the more moderate approach taken by other scholars in the legal realist movement. *See* Leiter, *supra* note 4, at 1148.

model of judging clearly wrong,[7] yet they also appear able to apply legal rules to facts, similarly disproving a purely realist model of judging.[8]

In this Article, we argue and attempt to demonstrate that neither the formalists nor the realists accurately describe the way judges make decisions, but that key insights from each form the core of a more accurate model.[9] We propose a blend of the two that we call the "intuitive-override" model of judging. Supported by contemporary psychological research on the human mind and by our own empirical evidence, this model posits that judges generally make intuitive decisions but sometimes override their intuition with deliberation. Less idealistic than the formalist model and less cynical than the realist model, our model is best described as "realistic formalism." The model is "realist" in the sense that it recognizes the important role of the judicial hunch and "formalist" in the sense that it recognizes the importance of deliberation in constraining the inevitable, but often undesirable, influence of intuition.[10]

Our model departs significantly from recent research on judicial decision making in two ways. First, most judicial scholars have studied appellate judges, particularly Supreme Court justices,[11] and their po-

---

7 Most scholars seem to agree that formalism does not accurately describe the way judges function. *See* LAURA KALMAN, LEGAL REALISM AT YALE, 1927–1960, at 229 (1986) ("We are all [legal] realists now."); *see also* David A. Skeel, Jr., *Corporate Anatomy Lessons*, 113 YALE L.J. 1519, 1542 (2004) (book review) ("When legal scholars announce that 'we are all legal realists now,' they usually mean that everyone now assumes that judicial opinions are more than simply the scientific application of existing law to each new set of facts.").

8 H.L.A. Hart called the realists' notion that judges are not bound by rules a "great exaggeration[ ]." H.L.A. HART, THE CONCEPT OF LAW 144 (1994). In rejecting the idea that deduction constrains judges, realism attracted much criticism. *See* Brian Leiter, *Legal Realism and Legal Positivism Reconsidered*, 111 ETHICS 278, 278–79, 300 (2001) (noting that Hart's critique of the realists in Chapter 7 of *The Concept of Law* turned Realism into a "jurisprudential joke, a tissue of philosophical confusions"). Even many of the realists conceded that judges were constrained by legal rules. *See* Joseph William Singer, *Legal Realism Now*, 76 CAL. L. REV. 465, 471 (1988) (book review) ("The most convincing legal realists argued that the reasoning demanded by judicial opinions substantially constrained judges.").

9 We do not consider the legal positivists, such as H.L.A. Hart, and their critics, such as Ronald Dworkin, as creating a theory of judicial decision making because they focus largely on providing a theory of law, not a descriptive account of judging itself. *See* Michael Steven Green, *Legal Realism as Theory of Law*, 46 WM. & MARY L. REV. 1915, 1917–18 (2005) (distinguishing between theories of law and theories of adjudication); Leiter, *supra* note 4, at 1140–44 (explaining that legal positivism is a theory of law, not a theory of adjudication).

10 *See* R. George Wright, *The Role of Intuition in Judicial Decisionmaking*, 42 HOUS. L. REV. 1381, 1420 (2006) ("Deciding judicial cases inescapably requires the exercise of intuition.").

11 *See, e.g.*, LAWRENCE BAUM, JUDGES AND THEIR AUDIENCES 4–5 (2006) ("[P]rimary attention to higher courts, especially the Supreme Court . . . . mirrors the subject matter of scholarship on judicial behavior."); JEFFREY A. SEGAL & HAROLD J. SPAETH, THE SUPREME

litical "attitudes" or ideology.[12] In contrast, our model arises from the study of trial judges. This is an important distinction because trial judges play a more prominent role in dispute resolution than do appellate judges.[13] Trial courts handle approximately 98% of the thirty-five million cases that the federal[14] and state[15] courts resolve each year. Moreover, trial court decisions are generally final because appeals are only available on limited bases,[16] occur infrequently,[17] and

---

COURT AND THE ATTITUDINAL MODEL 72–73 (1993). For an exception, see C. K. ROWLAND & ROBERT A. CARP, POLITICS & JUDGMENT IN FEDERAL DISTRICT COURTS 1 (1996).

12 *See* LEE EPSTEIN & JACK KNIGHT, THE CHOICES JUSTICES MAKE 9–12 (1998) (describing a political science approach to studying judges and the central role of judicial attitudes); SEGAL & SPAETH, *supra* note 11, at 72–73 (arguing that political attitudes are the primary influence on Supreme Court justices); HAROLD J. SPAETH, SUPREME COURT POLICY MAKING: EXPLANATION AND PREDICTION 113–18 (1979) (same).

13 In addition to presiding over jury trials, trial judges facilitate settlement, *see, e.g.*, Judith Resnik, *Managerial Judges*, 96 HARV. L. REV. 374, 376–77 (1982), resolve cases on motion, *see, e.g.*, Stephen B. Burbank, *Vanishing Trials and Summary Judgment in Federal Civil Cases: Drifting Toward Bethlehem or Gomorrah?*, 1 J. EMPIRICAL LEGAL STUD. 591, 592–93 (2004); Gillian K. Hadfield, *Where Have All the Trials Gone? Settlements, Nontrial Adjudications, and Statistical Artifacts in the Changing Disposition of Federal Civil Cases*, 1 J. EMPIRICAL LEGAL STUD. 705, 706–07 (2004), and decide more cases in bench trials than there are jury trials, *see, e.g.*, Brian J. Ostrom, Shauna M. Strickland & Paula L. Hannaford-Agor, *Examining Trial Trends in State Courts: 1976–2002*, 1 J. EMPIRICAL LEGAL STUD. 755, 773–74 (2004).

14 Of the approximately 2.3 million cases resolved in the federal courts during the 2006 term, 97% were resolved in the trial courts rather than the appellate courts. To derive this figure, we divided trial court resolutions (including bankruptcy court resolutions) by total resolutions during 2006: (1) 67,530 criminal case resolutions in district courts; (2) 281,220 civil case resolutions in district courts; (3) 1,889,778 resolutions in bankruptcy court; (4) 66,792 resolutions in courts of appeals (excluding the Federal Circuit); and (5) 82 resolutions in the Supreme Court. *See* U.S. COURTS, JUDICIAL CASELOAD INDICATORS, 12-MONTH PERIODS ENDING MARCH 31, 1997, 2002, 2005, AND 2006, http://www.uscourts.gov/caseload2006/front/mar06indicators.pdf (providing the District Court, Bankruptcy Court, and United States Court of Appeals data); PUB. INFO. OFFICE, 2006 YEAR-END REPORT ON THE FEDERAL JUDICIARY 9 (2007), http://www.supremecourtus.gov/publicinfo/year-end/2006year-endreport.pdf (providing the data for the Supreme Court).

15 Of the approximately 33 million court cases resolved in state courts during 2004 (the most recent year for which these data are available), 99% were resolved in the trial courts rather than the appellate courts. To derive this figure, we divided the total number of trial court resolutions in 2004 (32,137,043) by the total number of trial court resolutions plus the total appellate caseload (288,614) in 2004. For these data, see STATE COURT CASELOAD STATISTICS, 2005 tbls. 1, 10, http://www.ncsconline.org/D_Research/csp/2005_files/State%20Court%20Caseload%20Statistics%202005.pdf.

16 *See* ROWLAND & CARP, *supra* note 11, at 3; *see also* Maurice Rosenberg, *Standards of Review*, *in* RESTRUCTURING JUSTICE: THE INNOVATIONS OF THE NINTH CIRCUIT AND THE FUTURE OF THE FEDERAL COURTS 30, 31 (Arthur D. Hellman ed., 1990) (explaining that in many instances, "the court of appeals [is] obliged by established standards to affirm unless, for example, crucial fact findings were not merely in error but clearly so," and noting that "[d]iscretionary rulings [have] to be not merely incorrect, but abusive" to be reversed).

17 *See id.* at 8 ("[O]nly about 20 percent of all district court cases are appealed in any given year."); Theodore Eisenberg, *Appeal Rates and Outcomes in Tried and Nontried Cases: Further Exploration of Anti-Plaintiff Appellate Outcomes*, 1 J. EMPIRICAL LEGAL STUD. 659, 685 (2004) ("About 20 percent of cases with definitive trial court judgments generate appeals, with tried cases appealed at about twice the rate of nontried cases."). *But see* Chris Guthrie & Tracey E. George, *The Futility of Appeal: Disciplinary Insights Into the "Affirmance Effect" on*

seldom lead to reversal.[18]

Second, and relatedly, our model departs from prior research by identifying judicial *accuracy*, not judicial *activism*, as the most challenging issue facing the courts. As we demonstrate below, judges are predominantly intuitive decision makers, and intuitive judgments are often flawed. To be sure, intuition can lead to accurate decisions, as Malcolm Gladwell documents in his bestseller, *Blink*,[19] while deliberation can lead to error, as any court observer knows. But intuition is generally more likely than deliberation to lead judges astray.[20] We suspect this happens with some frequency, but even if it is uncommon, millions of litigants each year might be adversely affected by judicial overreliance on intuition. Therefore, the justice system should take steps to limit the impact of what we call "blinking on the bench."[21]

Eliminating all intuition from judicial decision making is both impossible and undesirable because it is an essential part of how the human brain functions.[22] Intuition is dangerous not because people rely on it but because they rely on it when it is inappropriate to do so. We propose that, where feasible, judges should use deliberation to check their intuition.

---

the United States Courts of Appeals, 32 FLA. ST. U. L. REV. 357, 361 fig.1b (2005) (reporting that the federal courts of appeal decide just under 30,000 cases on appeal each year). Dividing the 30,000 appeals by the 2.3 million resolutions yields an appeal percentage of only 1.3%.

18 *See* ROWLAND & CARP, *supra* note 11, at 8 ("[M]ost appeals are unsuccessful, and, as anticipated by the expanded fact freedom inherent in the evolution of fiduciary jurisprudence, the reversal rate is declining. For example, the reversal rate in 1960 was almost 25 percent; by 1990 it had declined to 16 percent. In combination, the low rates of appeal and reversal ensure that only a very small number of district courts' judgments will be reversed on appeal—about 3 percent(.20 x .16 = .032)."); Margaret A. Berger, *When, If Ever, Does Evidentiary Error Constitute Reversible Error?*, 25 LOY. L.A. L. REV. 893, 894 (1992) (finding only 30 trial verdicts in 1990 that were reversed for evidentiary error in the federal courts); Eisenberg, *supra* note 17, at 665 ("4.3 percent of the 2.1 million district court filings ended with an appellate court affirming the trial court and . . . 1.3 percent of such filings ended with an appellate court reversing the trial court. Thus, in rounded figures, 1 filing in 100 yields an appellate reversal. Simple computation reveals that about one-quarter of 1 percent of all cases filed from 1987 through 1995 led to an appellate court reversal of a *trial outcome* and that about one-half of 1 percent of such filed cases led to an appellate court affirmance of a trial outcome."); Guthrie & George, *supra* note 17, at 358 ("Affirmances are a defining feature of the courts of appeals . . . .").

19 *See* GLADWELL, *supra* note 1.

20 *See* ROBIN M. HOGARTH, EDUCATING INTUITION 224 (2001) ("[B]oth the tacit [intuitive] and the deliberate systems have their advantages and disadvantages. But they must be managed, and, for most people, this implies more active use of the deliberate system.").

21 Like Gladwell, we use the word "blink" to refer to the intuitive judgments made in the first moments after encountering a new situation. *See* GLADWELL, *supra* note 1, at 13–14 (discussing the "power of glance" derived from the first "two seconds" of experiencing something).

22 *See* HOGARTH, *supra* note 20, at 66 ("[T]he tacit system accounts for most mental activity.").

In Part I of this Article, we present our intuitive-override model of judging. In Part II, we present the results of our experimental research on judges. We provide tests of judges' general reasoning skills as well as their decision-making skills in legal contexts. Our results demonstrate that judges, like others, commonly make judgments intuitively, rather than reflectively, both generally and in legal contexts. Taken together, these data support our intuitive-override model.

Our model of judging raises two important questions for litigants, lawyers, judges, and the architects of the civil and criminal justice systems. First, which of the two decision-making approaches—intuitive or deliberative—is preferable? For reasons we explain below, we believe deliberative decision making is more likely than intuitive decision making to lead to just outcomes. Second, what might the justice system do to induce judges to decide matters more deliberatively and to produce more accurate outcomes? In Part III, we identify several concrete measures which the civil and criminal justice systems might implement to promote deliberative decision making.

I
THE INTUITIVE-OVERRIDE MODEL OF JUDGING

Our intuitive-override model of judging recognizes two distinct methods of judicial decision-making processes: intuitive and deliberative.

A. Dual-Process Models of Judgment

Philosophers have long distinguished between intuition and deduction.[23] Descartes, for example, claimed that "intuition and deduction" are the two processes "by means of which we [are able to] arrive at a knowledge of things."[24] Likewise, Pascal distinguished between the "intuitive" mind and the "geometric" mind.[25] "[W]ith the intuitive mind," he explained, "principles are in common use and before everybody's eyes. You have only to look, and no effort is necessary."[26] In the "geometric" mind, by contrast, "principles are obvious, but re-

---

23 *See, e.g.*, Seymour Epstein, *Integration of the Cognitive and the Psychodynamic Unconscious*, 49 AM. PSYCHOL. 709, 712 (1994) ("Awareness of a distinction between an experiential and a rational mode of processing information has a long history, predating psychology as a formal discipline."); Steven A. Sloman, *Two Systems of Reasoning, in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT 379, 380 (Thomas Gilovich, Dale Griffin & Daniel Kahneman eds., 2002) (observing that the distinction between intuition and deliberation "has not been missed by philosophers or psychologists" and "can be traced back to Aristotle").

24 RENÉ DESCARTES, RULES FOR THE DIRECTION OF THE NATURAL INTELLIGENCE 79 (George Heffernan ed. & trans., Rodopi 1998).

25 BLAISE PASCAL, PENSÉES 207 (Roger Ariew ed. & trans., Hackett Publ'g Co., Inc. 2005) (1670).

26 *Id.*

moved from ordinary use, so that we find it difficult to turn our head in that direction, for lack of habit."[27]

Building on these insights and on decades of research on judgment and choice, psychologists have proposed more than a dozen different dual-system or two-process models of cognition.[28] Although such models vary, they all distinguish between intuitive processes and deliberative processes.[29]

Intuitive processes, also called "System 1" processes,[30] "occur spontaneously and do not require or consume much attention."[31] They are "automatic, heuristic-based, and relatively undemanding of computational capacity."[32] Simply stated, they are "spontaneous, intuitive, effortless, and fast."[33] Emotional influences also tend to arise through System 1 processes.[34] Deliberative processes, also called "System 2" processes,[35] are "mental operations requiring effort, motivation, concentration, and the execution of learned rules."[36] Associated

---

27 *Id.*

28 *See, e.g.*, HOGARTH, *supra* note 20, at 21 ("The term *tacit system* is meant to encompass all processes that occur tacitly or automatically, that is, largely without use of conscious attention. . . . The term *deliberate system* is meant to encompass all processes that require effort, that is, attention and deliberation."); Epstein, *supra* note 23, at 711 tbl. 1, 715–19 (proposing a "cognitive-experiential self-theory" which includes an "experiential" system and a "rational" system); Jonathan St. B. T. Evans, *Heuristic and Analytic Processes in Reasoning*, 75 BRIT. J. PSYCHOL. 451 (1984) (proposing a dual system model including a "heuristic" system and an "analytic" system); Daniel Kahneman & Shane Frederick, *Representativeness Revisited: Attribute Substitution in Intuitive Judgment*, *in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT, *supra* note 23, at 49, 51–60 (making a dual-process claim and proposing their own model); Sloman, *supra* note 23, at 380–84 (proposing a dual system model with an "associative" system and a "rule-based" system); Keith E. Stanovich & Richard F. West, *Individual Differences in Reasoning: Implications for the Rationality Debate?*, *in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT, *supra* note 23, at 421, 436–38 (observing the burgeoning dual system models). *See generally* DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY (Shelly Chaiken & Yaacov Trope eds., 1999) (providing a collection of articles on various dual-process theories).

29 *See* Kahneman & Frederick, *supra* note 28, at 51.

30 Stanovich & West, *supra* note 28, at 436 (devising this label).

31 Shane Frederick, *Cognitive Reflection and Decision Making*, 19 J. ECON. PERSP. 25, 26 (2005); *see also* KENNETH R. HAMMOND, HUMAN JUDGMENT AND SOCIAL POLICY: IRREDUCIBLE UNCERTAINTY, INEVITABLE ERROR, UNAVOIDABLE INJUSTICE 60 (1996) ("The ordinary meaning of intuition signifies . . . a cognitive process that somehow produces an answer, solution, or idea without the use of a conscious, logically defensible, step-by-step process.").

32 Stanovich & West, *supra* note 28, at 436; *see also* HOGARTH, *supra* note 20, at 14 ("[T]he essence of intuition or intuitive responses is that they are reached with little apparent effort, and typically without conscious awareness. They involve little or no conscious deliberation."(emphasis omitted)).

33 Kahneman & Frederick, *supra* note 28, at 49 (referencing Tversky and Kahneman's earlier study of mathematically oriented p ychologists).

34 *See* Seymour Epstein & Rosemary Pacini, *Some Basic Issues Regarding Dual-Process Theories from the Perspective of Cognitive-Experiential Self-Theory*, *in* DUAL-PROCESS THEORIES IN SOCIAL PSYCHOLOGY, *supra* note 28, at 462, 469 ("[T]he experiential system [System 1] is emotionally driven.").

35 Stanovich & West, *supra* note 28, at 436 (adopting this label).

36 Frederick, *supra* note 31, at 26.

with "controlled processing,"[37] they are "deliberate, rule-governed, effortful, and slow."[38] Table 1 summarizes the differences between the two systems of processing.[39]

TABLE 1: A COMPARISON OF THE TWO SYSTEMS OF PROCESSING

Characteristic	**System 1 (Intuitive)**	**System 2 (Deliberative)**
Cognitive Style	Heuristic	Systematic
Cognitive Awareness	Low	High
Conscious Control	Low	High
Automaticity	High	Low
Speed	Fast	Slow
Reliability	Low	High
Effort	Low	High
Emotional Valence	High	Low

The relationship between the intuitive and the deliberative systems is complicated. Because intuition is automatic, quick, and easily invoked, it can easily dominate deliberation as decision makers simply rely on a quick, intuitive response or as intuition affects the judgments that follow.[40] Intuition can be surprisingly accurate, but sometimes good judgment will require purging the deliberative processes of intuition's influence.[41] Intuitive responses can also emerge from repetition of the same deliberative procedure.[42] Furthermore, some decisions might require shifting between the two systems.[43]

Our proposed dual-process model of judging, which is based on a general model developed by Daniel Kahneman and Shane Frederick,[44] posits that judges make initial intuitive judgments (System 1), which they might (or might not) override with deliberation (System

37 Stanovich & West, *supra* note 28, at 436.

38 Kahneman & Frederick, *supra* note 28, at 49 (referencing Tversky and Kahneman's earlier study of mathematically oriented psychologists).

39 We derived this table from Pat Croskerry, *The Theory and Practice of Clinical Decision-Making*, CANADIAN J. ANESTHESIA, June 1, 2005, at R1 & tbl. 1.

40 *See* HOGARTH, *supra* note 20, at 14 ("[E]ven if responses do involve conscious deliberation, the outcome may still depend on a set of initial intuitions.").

41 *See id.*

42 *See id.* at 22 ("[I]f deliberate-system actions are repeated over time, they can move to the domain of the tacit system.").

43 *See id.* ("[M]any cognitive activities do not rely on one system alone. . . . [I]nitial thoughts in a given situation may be the product of the tacit system at work, but these thoughts may be modified, amplified, or even rejected by the deliberate system."). For a model of judging that incorporates this kind of interaction, see Dan Simon, *A Third View of the Black Box: Cognitive Coherence in Legal Decision Making*, 71 U. CHI. L. REV. 511, 512–13 (2004) (proposing an alternate approach to classifying legal decision making as "coherence-based reasoning").

44 *See* Kahneman & Frederick, *supra* note 28, at 51.

2). As Kahneman and Frederick explain: "System 1 quickly proposes intuitive answers to judgment problems as they arise, and System 2 monitors the quality of these proposals, which it may endorse, correct, or override. The judgments that are eventually expressed are called *intuitive* if they retain the hypothesized initial proposal without much modification."[45] This model is similar to the one developed by psychologist Steven Sloman.[46] He uses the terms "associative" rather than System 1 or intuitive, and "rule-based" rather than System 2 or deliberative,[47] but he conceives of the judgment process in substantially the same way:

> Both systems seem to try, at least some of the time, to generate a response. The rule-based system can suppress the response of the associative system in the sense that it can overrule it. However, the associative system always has its opinion heard and, because of its speed and efficiency, often precedes and thus neutralizes the rule-based response.[48]

Our model, in short, views judges neither as the purely deductive decision makers envisioned by the formalists nor as the intuitive rationalizers envisioned by the early realists. Rather, it views judges as ordinary people who tend to make intuitive, System 1 decisions, but who can override their intuitive reactions with complex, deliberative thought.[49]

---

45 *Id.*

46 *See* Steven A. Sloman, *The Empirical Case for Two Systems of Reasoning*, 119 PSYCHOL. BULL. 3, 6 (1996); *see also* Stanovich & West, *supra* note 28, at 439 ("[O]ne of the functions of System 2 is to override some of the automatic contextualization provided by System 1."). This model also bears some resemblance to the more general "inferential correction" model of judgment proposed by Daniel Gilbert. *See* Daniel T. Gilbert, *Inferential Correction*, *in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT, *supra* note 23, at 167, 167 ("[O]ne of psychology's fundamental insights is that judgments are generally the products of nonconscious systems that operate quickly, on the basis of scant evidence, and in a routine manner, and then pass their hurried approximations to consciousness, which slowly and deliberately adjusts them.").

47 Sloman, *supra* note 46.

48 Sloman, *supra* note 23, at 391.

49 The convergence of psychologists on the notion that two separate systems of reasoning coexist in the human brain is remarkable. Psychologists Seymour Epstein, Steve Sloman, Dan Gilbert, Shelley Chaiken, and Daniel Kahneman come from diverse schools of thought, and yet all have settled on the same ideas about the dual-process models. Furthermore, these dual-process models find support from evolutionary psychology and neuropsychology. The brain consists of overlapping systems, each of which developed at a different point in the human evolutionary past. Contemporary neuropsychology supports the point. Studies of brain function reveal that people use different parts of the brain for different kinds of decisions. MRI scans of the brain reveal that the prefrontal lobes are most active when people are making deliberative decisions. *See* ELKHONON GOLDBERG, THE EXECUTIVE BRAIN: FRONTAL LOBES AND THE CIVILIZED MIND 69–70 (2001); Matthew D. Lieberman, Ruth Gaunt, Daniel T. Gilbert & Yaacov Trope, *Reflexion and Reflection: A Social Cognitive Neuroscience Approach to Attributional Inference*, 34 ADVANCES IN EXPERIMENTAL SOC. PSYCHOL. 199, 235 (2002). By contrast, the lateral temporal lobes, amygdala, and basal ganglia are active during intuitive, reflexive thought. *See* Daniel Schreiber & Marco

B. The Cognitive Reflection Test

The simplest and perhaps most powerful illustration of dual processing comes from Shane Frederick's "Cognitive Reflection Test" (CRT),[50] which lately has attracted much attention in both the academic and popular press.[51] The CRT is a three-item test designed to distinguish intuitive from deliberative processing.[52] More precisely, the CRT measures "cognitive reflection," which Frederick describes as "the ability or disposition to resist reporting the response that first comes to mind."[53] The CRT appears in its entirety in Figure 1 below:

FIGURE 1: COGNITIVE REFLECTION TEST[54]

(1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost?

_____cents

(2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

_____minutes

(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?

_____days

Each of the three CRT items has a correct answer that is easy to discern upon reflection, yet each also has an intuitive—but incorrect—answer that almost immediately comes to mind. Consider the first question. For many people, the answer that immediately jumps to mind is ten cents.[55] Though intuitive, this answer is wrong, as a bit of reflection shows. If the ball costs ten cents and the bat costs one dollar more, the bat must cost $1.10. Adding those two figures together, the total cost of the bat and ball would be $1.20, not $1.10. Therefore, the correct answer is five cents—the ball costs five cents, the bat costs $1.05, and together they cost $1.10.

For the second question, the answer that immediately jumps to mind is 100 minutes.[56] Though intuitive, this answer is also wrong. If

---

Iacoboni, *Evaluating Political Questions: Evaluation from Functional Brain Imaging* 7–8 (Sept. 1, 2004) (Harvard Univ. Weatherhead Ctr. for Int'l Affairs Paper), *available at* http://www.wcfia.harvard.edu/seminars/pegroup/Schreiber2004.pdf.

50 Frederick, *supra* note 31, at 26–28.

51 *See, e.g.*, Virginia Postrel, *Would You Take the Bird in the Hand, or a 75% Chance at the Two in the Bush*, N.Y. TIMES, Jan. 26, 2006, at C3; Peter Ranscombe, *The Three Questions that Will Show How Clever (or Not) You Really Are*, SCOTSMAN, July 16, 2005, http://news.scotsman.com/uk.cfm?id=1636262005.

52 *See* Frederick, *supra* note 31, at 27.

53 *Id.* at 35.

54 *Id.* at 27 fig.1.

55 *See id.* at 26–27.

56 *See id.* at 27.

five machines make five widgets in five minutes, then each machine makes one widget in that five-minute time period. Thus, it would take only five minutes for 100 machines to produce 100 widgets, just as 200 machines would make 200 widgets during that same period.[57]

The third question immediately invites an answer of twenty-four days, which is wrong.[58] The correct answer—obvious upon reflection—is forty-seven days. If the patch of lily pads doubles each day and covers the entire lake on the forty-eighth day, it must cover half the lake the day before.

The CRT items are simple in that "their solution is easily understood when explained, yet reaching the correct answer often requires the suppression of an erroneous answer that springs 'impulsively' to mind."[59] Most people, it turns out, are unable or unwilling to suppress that impulsive response.[60] In thirty-five separate studies involving 3428 respondents, Frederick found that subjects on average correctly answered 1.24 of the three items, although results varied across the subject pools.[61] For example, students at the University of Toledo obtained an average score of .57, while students at MIT obtained an average score of 2.18.[62] Among all of the subjects tested, only 17% answered all three questions correctly, while nearly twice that many (33%) answered all three questions incorrectly.[63]

Although the CRT consists of only three items, it correlates highly with the Wonderlic Personnel Test[64] (an intelligence test used by, among others, the National Football League[65]), the SAT,[66] and the

---

57 This problem implicitly assumes that each machine produces widgets at the same rate. Given the structure of the problem as well as the responses of subjects, we believe that subjects adopt that same assumption when confronted with this problem.

58 *See* Frederick, *supra* note 31, at 27.

59 *Id.*

60 *See id.* at 28, 29 tbl. 1.

61 *Id.*

62 *Id.* at 29.

63 *See id.* Among his other results, Frederick found that CRT scores correlate with time preferences. *See id.* at 30–32. Those who scored higher on the CRT were generally more willing to delay gratification and reward. *See id.* at 30–31. In addition, Frederick found that those who score higher on the CRT prefer the *New Yorker* to *People* magazine, whereas those who score lower on the CRT prefer *People* to the *New Yorker*. *See id.* at 39–40 n.15. Furthermore, Frederick found that CRT scores also correlate with risk preferences—those with high CRT scores are less likely to be influenced by the characterization of decision options as gains or losses from the status quo. *See id.* at 33 ("In the domain of gains, the high CRT group was more willing to gamble . . . . For items involving losses . . . , the high CRT group was *less* risk seeking . . . ."). Finally, Frederick found that men scored higher than women on the CRT—a result that he was not able to explain. *See id.* at 37–38.

64 *See id.* at 33–35. The correlation between the CRT and the Wonderlic Personality Test is .43. *Id.* at 35 tbl. 4.

65 *See id.* at 33–34.

66 *See id.* at 35. The correlation between the CRT and the SAT as a whole is 0.44. *Id.* at 35 tbl. 4.

ACT.[67] While the CRT acts as an abbreviated IQ test by measuring some component of intelligence, it would be a mistake to think of the CRT as simply that.[68] The CRT assesses a subset of what psychologists include in measures of intelligence—the capability and willingness to deliberate to solve a problem when intuition would lead one astray.[69]

The CRT illustrates the predominance of intuition over deliberation in three ways. First, adults perform poorly on the CRT, even though the questions are simple upon reflection.[70] Adults actually perform better on structurally similar problems that are *more* difficult, such as when the problem invites computation rather than impulse.[71] By way of illustration, Frederick explains that subjects "miss the 'bat and ball' problem far more often than they miss the 'banana and bagel' problem: 'A banana and a bagel cost 37 cents. The banana costs 13 cents more than the bagel. How much does the bagel cost?'"[72]

Second, among all potentially incorrect responses to the CRT questions, the intuitive answers identified above are most common.[73] By assessing "introspection, verbal reports and scribbles in the margin," Frederick found that even those subjects who responded correctly often considered the intuitive answer before selecting the correct answer.[74]

Third, subjects who selected the intuitive answers were more likely than those who answered correctly to indicate that the problems were easy.[75] In the bat-and-ball problem, Frederick found that subjects who provided the intuitive response predicted that 92% of people would solve the problem correctly, while subjects who responded correctly predicted that only 62% of people would do so.[76]

In sum, responses on the CRT provide evidence that ordinary adults possess dual-processing systems in which intuition tends to dominate, but can be overcome by deliberation. But what about

---

67 *See id.* at 35. The correlation between the CRT and the ACT is 0.46. *Id.* at 35 tbl. 4.

68 *See id.* at 33–37.

69 *See id.* at 35. Frederick is a bit coy about what the CRT attempts to measure, however. His assertion that it measures "the ability or disposition to resist reporting the response that first comes to mind" is somewhat vague. *Id.* at 35. The CRT might measure the decision maker's *willingness* to attend carefully to problems or *ability* to detect problems that produce intuitive but inaccurate answers. If the former, the CRT might shed little light on the behavior of judges, who might well be more willing to expend cognitive effort on the cases before them than on the abstract problems in the CRT. But the surprisingly high correlation between the CRT and other standard measures of intelligence suggests that the test measures the latter.

70 *See id.* at 27.

71 *See id.* at 28.

72 *Id.*

73 *See id.* at 27.

74 *Id.*

75 *See id.*

76 *Id.*

judges? On the one hand, it would be surprising if judges used markedly different decision-making processes; judges, after all, are human beings with the same cognitive machinery as everyone else. On the other hand, judges' education, intelligence, and on-the-job training as professional decision makers might distinguish them from most of the rest of the population.

II
TESTING THE MODEL

To explore whether judges make judgments consistent with our intuitive-override model, we measured their performance on the CRT and on a series of judicial decision-making problems. This work, which we describe in detail below, supports our theory that judges rely largely on intuition but sometimes override that intuition with deductive reasoning.

A. The CRT and Trial Judges

To explore whether judges behave like Frederick's subjects, we included the CRT in a five-item questionnaire we administered to 295 circuit court judges attending the Annual Business Meeting of the Florida Conference of Circuit Judges in Naples, Florida, on June 12, 2006. Florida's circuit court judges are the principal trial judges in the State. Of the 295 judges who returned surveys, 252 completed all of the items on the CRT,[77] meaning that nearly half of the circuit court judges in the Florida state courts completed the CRT.[78]

At this conference, we presented a plenary educational session to the judges entitled, "Judicial Decision Making." Although we do not have an exact count of the conference attendees, most of the judges participating in the conference attended our session. No other sessions ran at the same time. At the outset of our session, we distributed questionnaires to the judges in person and asked them to read and respond to each of the questions independently. The materials contained a cover page that indicated the name of the conference and provided the following instructions:

> Many of the points to be discussed at this session are best experienced directly. We therefore ask that before the session starts, you read and respond to each of the questions enclosed in this sur-

---

77 Seven judges answered none of the questions, and 288 judges answered at least one of the questions. In our analysis, we include only the 252 judges who answered all three questions.

78 *See* TOTAL STATEWIDE JUDGESHIPS: FISCAL YEAR 1972–73 TO FISCAL YEAR 2007–08 (2007), http://www.flcourts.org/gen_public/stats/bin/legauthorizedjudgesforweb.pdf (reporting that there were 564 circuit court judgeships in Florida during the 2005–06 fiscal year and 599 during the 2006–07 fiscal year).

> vey (although doing so is voluntary, of course). Please do so independently. Please do not discuss these materials while you are reviewing them, and please review the materials in the order presented. We shall collect these surveys before the discussion and present the results during this session.

At the beginning of the session, one of us (Wistrich) introduced himself and asked the judges to read and respond to the survey. He requested that they do so quietly, assured them that we were not collecting their names or other identifying information, and informed them that we would present a summary of the results at the end of the session. The judges appeared to take the questionnaires seriously. The room was silent during the administration of the questionnaires, which took approximately fifteen minutes.

Because we did not ask the judges to identify themselves, all responses were anonymous. We also informed the judges that participation was entirely voluntary. The final page of the questionnaires allowed the judges to limit the use of their answers to discussion during their particular conference, thereby excluding their answers from discussion in other contexts and from use in any publication. One judge exercised this option and we have excluded that judge's responses from our analysis. The CRT was the fourth item in our questionnaire. The last page of the questionnaire asked the judges to provide basic demographic information, including: their gender; the number of years of experience they have had as a judge; the major political party in the United States with which they most closely identify; the areas of judicial work in which they have had experience (civil, criminal, family, probate, or other); and their prior professional experiences.

At the top of the page on which we reproduced the CRT, we provided the same instructions Frederick provided to his subjects: "Below are several problems that vary in difficulty. Try to answer as many as you can."[79] Beneath the instructions, we reproduced the CRT, and beneath each of the CRT items we asked the judges to predict "[w]hat percentage of the judges in this room do you think will obtain the correct answer on this problem?"[80]

The judges obtained an average CRT score of 1.23 out of a possible 3.00. This score is slightly higher than the average that student subjects at Michigan achieved and slightly lower than the average student subjects at Harvard achieved.[81] Nearly one-third of the judges

---

79 *See* Frederick, *supra* note 31, at 28.

80 *See* stimulus materials on file with the authors.

81 Frederick, *supra* note 31, at 29 tbl. 1. Our results might overstate the judges' abilities on this problem because we excluded from the analysis the 43 judges who failed to answer all three questions. It is possible that these judges did not respond because they wanted to avoid answering questions they found difficult. Among the 36 judges who an-

(77 of 252 judges, or 30.6%) failed to answer a single question correctly; a similar number (78 out of 252 judges, or 31.0%) answered one question correctly; while less than one-quarter (60 out of 252 judges, or 23.8%) answered two questions correctly; and roughly one seventh (37 out of 252 judges, or 14.7%) answered all three questions correctly. That the judges produced results consistent with those of college students at highly selective universities suggests that the judges performed comparably to other well-educated adults.

TABLE 2: OVERALL CRT RESULTS: JUDGES COMPARED TO OTHERS[82]

Sample (n)	Mean	Percent with 0 correct	Percent with 1 correct	Percent with 2 correct	Percent with 3 correct
MIT (61)	2.18	7	16	30	48
Carnegie Mellon (746)	1.51	25	25	25	25
Harvard (51)	1.43	20	37	24	20
Florida judges (192)	1.23	31	31	24	15
Michigan/Ann Arbor (1267)	1.18	31	33	23	14
Bowling Green (52)	0.87	50	25	13	12
Michigan State (118)	0.79	49	29	16	6
Toledo (138)	0.57	64	21	10	5

The judges' performance improved as they progressed through the three questions, scoring 28.2%, 44.0%, and 50.4% correct on the first, second, and third questions, respectively. At first glance, that result seems odd because the second question is computationally more challenging than the first, yet more judges answered it correctly. Frederick's discussion of the CRT, however, predicts precisely this pattern because the second question seems more difficult than the first, which suggests to the test taker that reliance on intuition might be unwise.[83]

---

swered only one or two questions, the percentage of correct answers was 27.4% (17 out of 62), which is a little lower than the percentage correct among judges who answered all three questions (41%). The pattern of nonresponses, however, produced a dizzying array of combinations that does not clearly support any prediction as to how these judges would have performed if they had answered all three questions: 7 declined to answer any of the questions; 6 answered only the first question (5 of those answered it incorrectly); 4 answered only the second question (2 of these answered it incorrectly); 22 answered the first two questions (12 answered them both incorrectly, 2 answered them both correctly, 1 answered the first one correctly and the second incorrectly, and 7 answered the first one incorrectly and the second correctly); 3 answered only the first and third questions (2 answered them both incorrectly and 1 answered the first one incorrectly and the third one correctly); and 1 answered only the second and third questions (and answered the second question incorrectly and the third one correctly).

82 CRT scores for test takers other than Florida judges are taken from Frederick, *supra* note 31, at 29 tbl. 1.

83 *See* Frederick, *supra* note 31, at 27–28.

The judges who answered incorrectly tended to select the intuitively obvious, but inaccurate, responses. On the bat-and-ball question, 88.4% of those answering incorrectly (175 of 181 judges) provided the intuitive answer (10 cents). On the widget question, 57.4% of those answering incorrectly (81 of 141 judges) provided the intuitive answer (100 minutes). And on the lily pad question, 68% of those answering incorrectly (85 of 125 judges) provided the intuitive answer (24 days).[84]

TABLE 3: CRT RESULTS OF THE 252 JUDGES WHO ANSWERED ALL THREE QUESTIONS

	Percent accurate (n)	**Percent incorrect; giving the intuitive response (n)**	**Percent incorrect; giving any other response (n)**
Question 1	28.2% (71)	69.4% (175)	2.4% (6)
Question 2	44.0% (111)	32.1% (81)	23.8% (60)
Question 3	50.4% (127)	33.7% (85)	15.9% (40)

Finally, the judges who selected the intuitive, but incorrect, answers to a question were more likely than the judges who answered that question correctly to indicate that the question was easy. Table 4 reports these results.

TABLE 4: JUDGES' MEAN AND MEDIAN ESTIMATES OF THE PERCENTAGE OF OTHER JUDGES WHO WOULD GET EACH CRT ITEM CORRECT, BY ITEM AND BY JUDGES' ANSWERS (AMONG THE 252 JUDGES WHO ANSWERED ALL THREE QUESTIONS)

Question	**Judges Giving the Correct Answer**		**Judges Giving the Intuitive Wrong Answer**		**Judges Giving Unintuitive Wrong Answers**	
	Mean Percent (n)	Median Percent	Mean Percent (n)	Median Percent	Mean Percent (n)	Median Percent
1	64.6% (67)	75.0%	91.3% (168)	100.0%	53.3% (6)	50.0%
2	70.9% (104)	75.0%	80.9% (79)	95.0%	68.1% (57)	75.0%
3	67.6% (115)	75.0%	71.2% (82)	87.5%	36.2% (39)	25.0%

On the first question, the mean estimate of the percentage of judges who would answer the question correctly among those who selected the intuitive answer was 91.3%, while the mean estimate among judges who answered the question correctly was only 64.6%. This dif-

84 For our purposes, the CRT results are interesting primarily because our subjects are trial judges, and we are interested in understanding judicial decision making. These results should be of more general interest as well because as far as we know, our subjects are among the first nonstudent subjects to take the CRT and the first group of expert decision makers to take the test. *See id.* at 29 tbl. 1.

ference was statistically significant.[85] This gap diminished on the subsequent problems, but it did not disappear. On the second question, the mean estimate among judges who provided the intuitive answer was 80.9% versus 70.9% among those who provided the correct answer. This difference was also statistically significant.[86] And on the third question, the mean estimate of those who provided the intuitive answer was 71.2% versus 67.6% among those judges who answered correctly. This difference was not statistically significant.[87]

The demographic characteristics of the judges (gender, years of experience, and political party) did not correlate with CRT scores. The 185 male judges scored a mean of 1.28, while the 58 female judges scored a mean of 1.21. This difference was not statistically significant.[88] The correlation coefficient between years of experience and CRT score was .06, which was also not significant.[89] Finally, the 88 judges reporting that they most closely identified with the Democratic Party scored a mean of 1.16, while the 121 judges who identified with the Republican Party scored a mean of 1.24. This difference also was not significant.[90]

Collectively, these results suggest that judges tended to favor intuitive rather than deliberative faculties. First, the judges had trouble with the CRT, even though the CRT questions are not difficult. Most

---

85 We performed this analysis with a one-way ANOVA with three levels corresponding to the three possible answers (correct, wrong-intuitive, and wrong-other). The ANOVA revealed significant difference among the groups. $F(2, 238) = 37.99$, $p < .001$. Post hoc analysis using Tukey's familywise errors revealed that all three groups were significantly different from each other.

86 We performed this analysis with a one-way ANOVA with three levels, corresponding to the three possible answers (correct, wrong-intuitive, and wrong-other). The ANOVA revealed significant difference among the groups. $F(2, 237) = 4.63$, $p = .011$. Post hoc analysis using Tukey's familywise errors revealed that all the judges who provided the intuitive answer differed significantly from the judges who provided the correct answer and from the judges who provided other wrong answers. The judges who provided the correct answers did not differ significantly from the judges who provided nonintuitive wrong answers.

87 We performed this analysis with a one-way ANOVA with three levels, corresponding to the three possible answers (correct, wrong-intuitive, and wrong-other). The ANOVA revealed significant difference among the groups. $F(2, 233) = 22.08$, $p < .001$. Post hoc analysis using Tukey's familywise errors revealed that the significant result here was driven entirely by the judges who provided nonintuitive wrong answers. The judges who answered correctly did not provide significantly different estimates from judges who provided intuitive answers.

88 $t(241) = 1.02$, $p = .31$. Note that 9 of the 252 judges who answered all three CRT questions failed to identify their gender. In his work Frederick found a greater tendency for men to score higher than women. *See* Frederick, *supra* note 31, at 37–38.

89 The ordered logit regression of CRT score on years of experience did not produce a significant relationship. $F(1,236) = 0.79$, $p = 0.375$. Note that 14 of the 252 judges who answered all three CRT questions failed to identify the number of years of experience they had had as a judge.

90 $t(207) = 0.55$, $p = .58$. Note that 43 of the 252 judges who answered all three CRT questions failed to identify their political party.

of the judges answered most of the questions wrong. Second, when the judges erred, they generally chose the intuitive answer. Third, those judges who selected the intuitive answer indicated that the problem was easier than those judges who suppressed their intuition and provided a deliberative answer.

Despite the predominantly intuitive responses the judges provided, the judges also demonstrated that they can override their intuitive responses. Roughly two-thirds of the participants answered one or more of the CRT problems deliberatively,[91] and roughly one-seventh of the judges answered all three deliberatively.[92] This put the judges in good company, obtaining scores relatively close to those of Harvard undergraduates.[93] Nevertheless, the profession of judging clearly does not attract exclusively deductive, System 2 thinkers.

To be sure, the judges undoubtedly exercise greater care when ruling in court than when responding to CRT questions at an educational conference. Moreover, we did not provide any tangible incentive to the judges, in contrast to Frederick, who paid subjects in most of his studies $8 to complete a lengthy questionnaire.[94] We do not believe that either of these factors undermine our results, however. First, although Frederick paid his subjects, he paid them only for participation, not for correct responses.[95] Second, even though we did not provide tangible incentives, we informed the judges that we intended to share their collective results with the group, which probably induced many judges to try to solve these problems correctly. Third, the judges appeared to take our questionnaires quite seriously, completing them in a quiet and focused manner. Fourth, although this is admittedly anecdotal, the judges appeared deeply interested in learning the results of the CRT; in subsequent break-out sessions they wanted to discuss these results more than any of the other materials we presented. Fifth, incentives do not always improve judgment and decision making.[96] Finally, in contrast to Frederick, who embedded the CRT in a questionnaire designed to take 45 minutes to complete, we included the CRT in a five-item questionnaire designed to take only 15 minutes.[97] Thus, the judges participating in our study were likely less mentally taxed, under less time pressure, and more attentive than were the subjects participating in Frederick's studies. We cannot

---

91 *See supra* Part I.B.

92 *See id.*

93 *See* Frederick, *supra* note 31, at 29 tbl. 1.

94 *See id.* at 28.

95 *See id.*

96 *See, e.g.*, Colin F. Camerer & Robin M. Hogarth, *The Effects of Financial Incentives in Experiments: A Review and Capital-Labor Production Framework*, 19 J. RISK & UNCERTAINTY 7 (1999) (reporting mixed evidence of incentive effects).

97 *See* Frederick, *supra* note 31, at 28.

eliminate the possibility that more highly motivated judges might perform better on the CRT tasks, but we suspect that even highly motivated judges would produce similar results.

B. Studies of Judicial Decision Making

The CRT results suggest that judges tend to make *ordinary* judgments intuitively. The fact that judges demonstrate a predominantly intuitive approach to the CRT questions does not necessarily mean that they make predominantly intuitive judgments *as judges*. In recent years, however, we have conducted several studies involving hundreds of federal and state trial judges around the nation, and we have found that judges commonly encounter stimuli on the job that induce intuitive reactions, though they occasionally demonstrate an ability to override those intuitive responses.[98] We do not intend to catalog our results here. But to illustrate the claim we make in this paper, we describe three examples of how judges react to cues that trigger intuitive decision making. Specifically, we explore how judges respond to numeric anchors, evaluate statistical evidence, and assess conduct after learning an outcome associated with that conduct. Collectively, this research shows that judges tend to make decisions in a largely intuitive way.

1. *Intuitive Judging—Anchoring*

The first example of intuitive judicial decision making arises from studies of a phenomenon that psychologists call "anchoring."[99] When making numeric estimates, people commonly rely on the initial value available to them.[100] This initial value provides a starting point that "anchors" the subsequent estimation process.[101] People generally adjust away from the anchor, but typically fail to adjust sufficiently, thereby giving the anchor greater influence on the final estimate than it should have.[102] In short, "the number that starts the generation of a

---

98 *See* Chris Guthrie, Jeffrey J. Rachlinski & Andrew J. Wistrich, *Inside the Judicial Mind*, 86 CORNELL L. REV. 777, 779–80 (2001) (reporting experimental evidence showing that judges are susceptible to heuristics and biases when making judgments); Jeffrey J. Rachlinski, Chris Guthrie & Andrew J. Wistrich, *Inside the Bankruptcy Judge's Mind*, 86 B.U. L. REV. 1227, 1229–30 (2006) (exploring whether specialized, bankruptcy judges are similarly susceptible to heuristics and biases); Andrew J. Wistrich, Chris Guthrie & Jeffrey J. Rachlinski, *Can Judges Ignore Inadmissible Information? The Difficulty of Deliberately Disregarding*, 153 U. PA. L. REV. 1251, 1258–59 (2005) (reporting experimental evidence showing that judges have difficulty deliberately disregarding relevant but inadmissible evidence when making merits decisions).

99 *See* Amos Tversky & Daniel Kahneman, *Judgment Under Uncertainty: Heuristics and Biases*, 185 SCI. 1124, 1128 (1974) (identifying anchoring).

100 *See id.*

101 *See id.* ("[D]ifferent starting points yield different estimates, which are biased toward the initial values.").

102 *See id.*

judgment exerts a stronger impact than do subsequent pieces of numeric information."[103]

We have found that anchors trigger intuitive judicial decision making.[104] In one study, we demonstrated that a demand made at a prehearing settlement conference anchored judges' assessments of the appropriate amount of damages to award.[105] In that study, we randomly assigned the participating judges to either a control group or an anchor group and presented them with a lengthy vignette describing a civil case in which the plaintiff had suffered multiple injuries in a car accident caused by a negligent truck driver:

> Imagine that you are presiding over an automobile accident case in which the parties have agreed to a bench trial. The plaintiff is a 31-year-old male schoolteacher and the defendant is a large package-delivery service. The plaintiff was sideswiped by a truck driven erratically by one of the defendant's drivers. As a result of the accident, the plaintiff broke three ribs and severely injured his right arm. He spent a week in the hospital, and missed six weeks of work. The injuries to his right arm were so severe as to require amputation. (He was right-handed.)[106]

We informed the judges that they had presided over an unsuccessful settlement conference in this case. The judges in the control group learned that the plaintiff's lawyer had told them at the settlement conference that the plaintiff "was intent upon collecting a significant monetary payment." The judges in the high anchor group learned that the plaintiff's lawyer had demanded $10 million. We asked the judges in both groups to indicate the amount of compensatory damages they would award the plaintiff.

The $10 million anchor influenced the judges. Judges in the control group awarded a mean amount of $808,000 and a median amount of $700,000, while judges in the anchor group awarded a much larger mean of $2,210,000 and median of $1 million.[107] Table 5 shows the impact the anchor had on their judgment.

---

103 Fritz Strack & Thomas Mussweiler, *Heuristic Strategies for Estimation Under Uncertainty: The Enigmatic Case of Anchoring*, *in* FOUNDATIONS OF SOCIAL COGNITION 79, 80 (Galen V. Bodenhausen & Alan J. Lambert eds., 2003).

104 *See* Guthrie et al., *supra* note 98, at 790–94; Wistrich et al., *supra* note 98, at 1286–93.

105 *See* Wistrich et al., *supra* note 98, at 1286–93.

106 *Id.* at 1332.

107 *Id.* at 1290.

TABLE 5: ANCHORING STUDY OF SETTLEMENT DEMAND (IN $1,000S)[108]

	Mean	1st Quartile	Median	3rd Quartile
Control (37)	808	325	700	1000
Anchor (38)	2210	575	1000	3000

In another study, we tested whether a motion to dismiss would also affect judges' damage awards.[109] We presented participating judges with a similar fact pattern and asked judges in the control group, "[H]ow much would you award the plaintiff in compensatory damages?"[110] We gave the judges in the anchor group the same background information, but also told them that "[t]he defendant has moved for dismissal of the case, arguing that it does not meet the jurisdictional minimum for a diversity case of $75,000." We asked these judges to rule on the motion, and then asked them, "If you deny the motion, how much would you award the plaintiff in compensatory damages?" Because the plaintiff clearly had incurred damages greater than $75,000, we viewed the motion as meritless, as did all but two of the judges.[111] Nonetheless, the $75,000 jurisdictional minimum served as an anchor and resulted in lower damage awards from those judges exposed to it. The judges who had not ruled on the motion awarded the plaintiff an average of $1,249,000 (and a median of $1 million), while those judges who ruled on the motion to dismiss awarded the plaintiff an average of $882,000 (and a median of $882,000).[112] Thus, the $75,000 jurisdictional minimum anchored the judges' assessments, as they awarded roughly $350,000 (or nearly 30%) less on average.

TABLE 6: ANCHORING STUDY OF MOTION TO DISMISS (IN $1000S)[113]

	Mean	1st Quartile	Median	3rd Quartile
Control (66)	1249	500	1000	1925
Anchor (50)	882	288	882	1000

Both anchoring studies suggest that the anchors had a powerful influence on judgment. This was true both when the anchor bore essentially no relation to the magnitude of the claim and when the judges knew full well that they were supposed to ignore the anchor. In both cases, the anchor triggered intuitive, automatic processing that the judges were unable to override.

---

108 *See id.*

109 *See* Guthrie et al., *supra* note 98, at 790.

110 *Id.* at 790–91.

111 *See id.* at 791.

112 *Id.* at 791–92. The difference in response rate between the control and anchor groups was statistically significant. *See id.* at 791 n.69.

113 *Id.* at 791–92.

2. *Intuitive Judging—Statistical Inferences*

The second example of intuitive judicial decision making arises from studies of what psychologists call the "representativeness" heuristic.[114] When people rely on the representativeness heuristic, they tend to undervalue statistical information, which can lead to notable decision errors.[115] For example, people tend to discount information about the frequency with which the underlying category occurs, a phenomenon known as "base rate" neglect.[116] In one illustrative study, researchers asked college students to indicate whether a person described as being "of high intelligence, although lacking . . . creativity" who "has a high need for order and clarity" and whose "writing is rather dull" and who seems to have "little sympathy for other people and does not enjoy interacting with others" was a student in either computer science or in humanities and education.[117] Although the participants knew that three times as many graduate students studied humanities and education as studied computer science, they tended to guess that the student was in computer science.[118] Notwithstanding the high relevance of base-rate statistics, people discount their probative value in favor of impressionistic and intuitive reactions to the representativeness of the information.[119]

To test whether judges would rely on their intuitive assessments rather than statistical information when presented with a case, we gave a group of federal magistrate judges the following problem, based on the classic English case, *Byrne v. Boadle*:[120]

---

114 *See* Daniel Kahneman & Amos Tversky, *Subjective Probability: A Judgment of Representativeness*, 3 COGNITIVE PSYCHOL. 430, 430 (1972) (first defining the "representativeness" heuristic); *see also* Daniel Kahneman & Amos Tversky, *On the Psychology of Prediction*, 80 PSYCHOL. REV. 237, 237 (1973) (expanding on their earlier treatment of the representativeness heuristic) [hereinafter *Prediction*]; Amos Tversky & Daniel Kahneman, *Belief in the Law of Small Numbers*, 76 PSYCHOL. BULL. 105, 105 (1971) (exploring the human tendency to treat a sample as more representative of a population than is justified by probability theory) [hereinafter *Small Numbers*]; Amos Tversky & Daniel Kahneman, *Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment*, 90 PSYCHOL. REV. 293, 293 (1983) (exploring the "conjunction fallacy," a manifestation of the representativeness heuristic); Amos Tversky & Daniel Kahneman, *Judgments of and by Representativeness*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES 84, 84–85 (Daniel Kahneman, Paul Slovic & Amos Tversky eds., 1982) (synthesizing the authors' prior work on representativeness); Tversky & Kahneman, *supra* note 99, at 1124–27 (explaining various decision errors caused by the representativeness heuristic).

115 *See* Tversky & Kahneman, *supra* note 99, at 1124–27.

116 *See Prediction*, *supra* note 114, at 239; *see also* Tversky & Kahneman, *supra* note 99, at 1124.

117 *Prediction*, *supra* note 114, at 238.

118 *See id.* at 239.

119 *See id. But see* Jonathan J. Koehler, *The Base Rate Fallacy Reconsidered: Descriptive, Normative, and Methodological Challenges*, 19 BEHAV. & BRAIN SCI. 1, 1–2 (1996) (arguing that the proponents of the representativeness heuristic have overstated the extent to which people actually neglect base rates).

120 159 Eng. Rep. 299 (1863); *see* Guthrie et al., *supra* note 98, at 808.

> The plaintiff was passing by a warehouse owned by the defendant when he was struck by a barrel, resulting in severe injuries. At the time, the barrel was in the final stages of being hoisted from the ground and loaded into the warehouse. The defendant's employees are not sure how the barrel broke loose and fell, but they agree that either the barrel was negligently secured or the rope was faulty. Government safety inspectors conducted an investigation of the warehouse and determined that in this warehouse: (1) when barrels are negligently secured, there is a 90% chance that they will break loose; (2) when barrels are safely secured, they break loose only 1% of the time; (3) workers negligently secure barrels only 1 in 1,000 times.[121]

We then asked: "'Given these facts, how likely is it that the barrel that hit the plaintiff fell due to the negligence of one of the workers?'"[122] The materials then asked the judges to answer by choosing one of four probability ranges: 0–25%, 26–50%, 51–75%, or 76–100%.

When presented with a problem like this one, most people rely on their intuition[123]—the accident sounds like it was the product of negligence, so intuition would suggest negligence must have caused it. The subjects largely treat the 90% figure as the likelihood that the accident was the product of negligence, thereby converting the true meaning of the 90% statistic (the likelihood of injury given negligence) into its inverse (the likelihood of negligence given injury).[124] A deductive approach reveals that the actual probability that the defendant was negligent is only 8.3%.[125]

Most of the judges who assessed our problem answered it incorrectly.[126] In fact, only about 40% answered correctly and selected the low range as the actual probability that the accident was the result of negligence.[127] Much like the CRT results, the most common wrong answer (also selected by roughly 40% of the judges) was the intuitive

---

121 Guthrie et al., *supra* note 98, at 808.

122 *Id.*

123 *See supra* notes 114–119 and accompanying text (describing how people often react intuitively based on similarity information rather than analyzing deliberatively using relevant statistical information like base rates).

124 *See* Guthrie et al., *supra* note 98, at 808–09.

125 *See id.* at 809 ("Because the defendant is negligent .1% of the time and is 90% likely to cause an injury under these circumstances, the probability that a victim would be injured by the defendant's negligence is .09% (and the probability that the defendant is negligent but causes no injury is .01%). Because the defendant is not negligent 99.9% of the time and is 1% likely to cause an injury under these circumstances, the probability that on any given occasion a victim would be injured even though the defendant took reasonable care is 0.999% (and the probability that the defendant is not negligent and causes no injury is 98.901%). As a result, the conditional probability that the defendant is negligent given that the plaintiff is injured equals .090% divided by 1.089%, or 8.3%.").

126 *See id.*

127 *See id.* at 809.

response that the accident was more than 75% likely to have been the product of negligence.[128]

Compared to other people who have evaluated similar statistical problems, the judges we studied performed well. Fewer than 20% of doctors facing a nearly identical problem in a medical context chose the correct answer.[129] Thus, although many of the judges responded intuitively, many others responded deliberatively such that the overall relative performance of judges was admirable.

3. *Intuitive Judging—Hindsight*

The third example of intuitive judicial decision making comes from studies of the "hindsight bias."[130] The hindsight bias is the well-documented tendency to overestimate the predictability of past events.[131] The bias arises from an intuitive sense that the outcome that actually happened must have been inevitable. People allow their knowledge to influence their sense of what would have been predictable.[132]

Because judges usually evaluate events after the fact, they are vulnerable to the hindsight bias.[133] To explore whether judges would be prone to the hindsight bias, we gave participating judges a hypotheti-

---

128 *See id.* at 810.

129 *See* Ward Casscells, Arno Schoenberger & Thomas B. Graboys, *Interpretation by Physicians of Clinical Laboratory Results*, 299 NEW ENG. J. MED. 999, 999–1000 (1978).

130 *See generally* Baruch Fischhoff, *Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty*, 1 J. EXPERIMENTAL PSYCHOL.: HUM. PERCEPTION & PERFORMANCE 288, 288–99 (1975) (documenting the effects of the hindsight bias).

131 *See id.*; *see also* Baruch Fischhoff, *For Those Condemned to Study the Past: Heuristics and Biases in Hindsight*, *in* JUDGMENT UNDER UNCERTAINTY: HEURISTICS AND BIASES, *supra* note 115, at 335, 341–43.

132 *See* Scott A. Hawkins & Reid Hastie, *Hindsight: Biased Judgments of Past Events After the Outcomes Are Known*, 107 PSYCHOL. BULL. 311, 312–13 (1990). To be sure, the hindsight bias has a deliberative component as well. Knowledge of the outcome seems to affect how people interpret antecedent circumstances, which seems more deliberative than intuitive. *See id.* As we noted above, the relationship between intuition and deliberation can be complicated.

133 *Cf.* Kim A. Kamin & Jeffrey J. Rachlinski, *Ex Post ≠ Ex Ante: Determining Liability in Hindsight*, 19 LAW & HUM. BEHAV. 89, 93, 99 (1995) (finding that the hindsight bias influenced jurors' liability determinations in a simulated negligence case). In addition to influencing determinations of negligence, the hindsight bias likely influences claims of ineffective assistance of counsel (decisions a lawyer makes in the course of representing a criminal defendant can seem less competent after the defendant has been convicted), the levying of sanctions under Rule 11 of the Federal Rules of Civil Procedure (a motion or allegation seems less meritorious after a court rejects it), and assessments of the liability of corporate officers charged with making false predictions about their company's performance (which can look like fraud after the predictions fail to come true). *See* Jeffrey J. Rachlinski, *A Positive Psychological Theory of Judging in Hindsight*, 65 U. CHI. L. REV. 571, 602–24 (1998). On the relationship between hindsight bias and securities fraud in particular, see Mitu Gulati, Jeffrey J. Rachlinski & Donald C. Langevoort, *Fraud by Hindsight*, 98 NW. U. L. REV. 773, 824–25 (2004).

cal fact pattern based on an actual case, labeled "Likely Outcome of Appeal":[134]

> In 1991, a state prisoner filed a *pro se* Section 1983 action in Federal District Court against the Director of the Department of Criminal Justice in his state, asserting, among other things, that the prison had provided him with negligent medical treatment in violation of Section 1983. The district court dismissed his complaint on the ground that the provision of negligent medical care does not violate Section 1983. The district court further found that the plaintiff knew his claims were not actionable because he had made similar claims several years earlier in a case that had been dismissed by the court. Thus, the district court sanctioned the plaintiff pursuant to Rule 11, ordering him to obtain the permission of the Chief Judge in the district before filing any more claims. The plaintiff appealed the district court's decision.[135]

Each judge randomly received one of three conditions: "Affirmed," "Vacated," or "Lesser Sanction." Judges in each condition learned that a different outcome had been obtained on appeal:

- 'Lesser Sanction': 'The court of appeals ruled that the district court had abused its discretion under Rule 11 and remanded the case for imposition of a less onerous Rule 11 sanction against the plaintiff.'
- 'Affirmed': 'The court of appeals affirmed the district court's decision to impose this Rule 11 sanction on the plaintiff.'
- 'Vacated': 'The court of appeals found that the district court had abused its discretion and vacated the Rule 11 sanction against the plaintiff.'

We asked the judges in each group the following: "In light of the facts of the case, as described in the passage above, which of the following possible outcomes of the appeal was most likely to have occurred (assume that the three outcomes below are the only possible ones)?"

The judges' responses were influenced by learning the outcome on appeal. Among the judges informed that the court of appeals had remanded for a lesser sanction, 38.6% asserted that they would have predicted that outcome, as compared to 7.4% and 20.4% of the judges whom we informed that the court of appeals had affirmed and vacated the sanction, respectively.[136] Among judges whom we informed that the court of appeals had affirmed, 81.5% indicated that they would have predicted that result, as compared to only 40.4% and 27.8% of judges whom we informed that the court of appeals had or-

---

134 Guthrie et al., *supra* note 98, at 801.

135 *Id.*

136 *Id.* at 803 tbl. 2.

dered a lesser sanction and vacated, respectively. Finally, among judges whom we informed that the court of appeals had vacated, 51.9% indicated that they would have predicted that result, as compared to only 21.1% and 11.1% of judges whom we informed that the court of appeals had ordered a lesser sanction and affirmed, respectively. The sum of the percentage of judges in each condition who identified the outcome they were given as the "most likely to have occurred" was 172%; if knowing the outcome had not influenced the judges, this sum would have been only 100%. Learning an outcome clearly influenced the judges' ex post assessments of the ex ante likelihood of various possible outcomes. The intuitive notion that the past was predictable prevailed.

TABLE 7: HINDSIGHT BIAS APPELLATE OUTCOME STUDY[137]

	Percent selecting "lesser" as outcome	Percent selecting "affirmed" as outcome	Percent selecting "vacate" as outcome
Lesser Group (57)	38.6	40.4	21.1
Affirm Group (54)	7.4	81.5	11.1
Vacate (54)	20.4	27.8	51.9

In another study, however, we found that judges are sometimes capable of resisting the hindsight bias.[138] Using a problem involving an assessment of probable cause, we randomly assigned judges to either a foresight group or a hindsight group.[139] We asked the judges in the foresight group whether they would grant a search warrant under circumstances we described, and we asked the judges in the hindsight group to rule on the admissibility of evidence gathered without a warrant under the same circumstances.[140] We then compared their responses.

The materials we provided to judges in both groups provided the same core set of facts.[141] The materials stated that a police officer was on patrol outside a rock concert. The officer saw a well-dressed, nervous-looking man exit a BMW and fiddle with something in the trunk before he entered the concert. A half hour later, the officer noticed that one of the BMW's windows was down. Concerned that someone might burglarize the car, he approached to close the window. Upon reaching the car, he "smelled something that he believed, based on a demonstration at a training session several years earlier, to be burnt methamphetamine. He looked inside the car and didn't see any

137 *Id.* at 803 tbl. 2.

138 *See* Wistrich et al., *supra* note 98, at 1251–52.

139 *See id.* at 1314–15.

140 *See id.* at 1315.

141 *See id.* (describing the materials in this study).

drugs, but he did notice some Visine, a local map, and a couple of empty beer cans."

The judges assigned to the foresight group learned that the police officer believed that probable cause existed and called to request a telephonic warrant to search the trunk of the car. We asked these judges to indicate whether they would issue the telephonic warrant. The judges assigned to the hindsight group learned that the police officer conducted a warrantless search of the trunk and found ten pounds of methamphetamine, other drug paraphernalia, and a recently fired gun that had been used earlier in the day to murder a drug dealer across town. The police officer arrested the car owner, who was subsequently prosecuted. During his prosecution, his defense attorney moved to suppress the evidence, arguing that the officer did not have probable cause to search the trunk. We asked these judges to indicate whether they would admit this evidence.

The responses of the judges in the foresight and hindsight groups were statistically indistinguishable. In the foresight condition, 23.9% of the judges indicated that there was probable cause to issue the warrant, while in the hindsight condition, 27.7% of the judges found probable cause to conduct the search and ruled the evidence admissible.[142]

Although the hindsight bias affected judges in the "appeals" problem, judges demonstrated resistance to the bias in the "probable cause" problem. The highly intricate, rule-bound nature of Fourth Amendment jurisprudence that guides probable cause determinations might have facilitated the deliberative, System 2 approach. The intricacy of this area of law signals to judges that intuition might be inconsistent with the governing law and therefore that they will need to think carefully through the rules created by the appellate courts. Just as the apparent intricacies of the "banana and bagel" problem—in contrast to the "bat and ball" problem—induce decision makers to deliberate, so too might the complexities of the rules governing probable cause.[143] In contrast, the appeals problem presents no intricate collection of rules and case law to signal judges that they should not rely on their intuition.

4. *Summary of Intuitive Judging Studies*

These results suggest that judges rely heavily on their intuitive faculties not only when they confront generic problems like the problems included in the CRT, but also when they face the kinds of problems they generally see on the bench. When awarding damages,

---

142 *Id.*

143 *See supra* note 72 and accompanying text.

assessing liability based on statistical evidence, and predicting outcomes on appeal, judges seem inclined to make intuitive judgments. They are also vulnerable to such distractions as absurd settlement demands, unrelated numeric caps, and vivid fact patterns.

But our studies also show that judges can sometimes overcome their intuitive reactions and make deliberative decisions. Our study of probable cause revealed that hindsight bias had no effect in certain contexts. The results of this study were surprising considering the body of research that we and others[144] have conducted, which would seem to predict the opposite result.[145] Together, this work suggests that judges are inclined, at least when presented with certain stimuli, to make intuitive decisions, but that they have the capacity to override intuition with deliberative thinking.

How can we account for the prominent role intuition played in some of our studies and not in others? On this question, we believe the results of our two hindsight bias problems are particularly illuminating.[146] First, trial judges have far more discretion on their choice of sanctions—at issue in the first hindsight bias problem reported above—than their determination of probable cause—at issue in the second hindsight bias problem reported above. The latter area, in contrast to the former, includes a web of complex rules familiar to most trial judges. This web of rules might enable trial judges to avoid the hindsight bias. Second, the questions we asked the judges varied in a subtle way. In the first problem reported above, we did not ask them to assess whether the sanction itself was appropriate, but rather

---

144 *See, e.g.*, Theodore Eisenberg, *Differing Perceptions of Attorney Fees in Bankruptcy Cases*, 72 WASH. U. L.Q. 979, 982–87 (1994) (reporting evidence suggesting that judges are prone to the self-serving or egocentric bias); W. Kip Viscusi, *How Do Judges Think About Risk?*, 1 AMER. L. & ECON. REV. 26, 29 (1999) (finding that "judges exhibit a variety of biases"); *see also* Rachlinski, *supra* note 133, at 580 ("Virtually every study on judging in hindsight has concluded that events seem more predictable than they actually are.").

145 *But see* Rachlinski et al., *supra* note 98, at 1256–57 (finding, despite evidence of susceptibility to anchoring and framing, that bankruptcy judges appeared uninfluenced by omission bias and some emotional factors); Viscusi, *supra* note 144, at 46–55 (finding, among judges attending a law and economics conference, little evidence of hindsight bias).

146 Even in the anchoring and hindsight study of appeals, the fact that we generally observed statistically significant differences between the control group judges and experimental group judges does not mean that every judge made intuitive decisions. Take, for example, the anchoring studies described above. *See supra* Part II.B.1. In those studies, we found that the group of judges exposed to an anchor responded differently from the group of judges not exposed to that anchor. *See id.* This does not mean that every judge in the anchor group made an intuitive judgment; indeed, some judges in the anchor group resisted the influence of the anchor and responded similarly to the judges in the control group (just like some judges who took the CRT overcame their intuitive reactions on one, two, or all of three of the problems). Our results only show that, as a group, the judges were heavily influenced by their intuition—they do not tell us which judges were influenced and by how much.

to identify the likely outcome on appeal. Asking them for a prediction rather than a legal ruling might have prompted them to make an intuitive estimate. In the probable cause problem, by contrast, we asked the judges to rule on the admissibility of the evidence or to grant or deny a warrant; as a result, we might have prompted them to think carefully about a familiar body of case law. So prompted, they were able to override their intuitive reactions with deliberation.

III
IMPLICATIONS FOR THE JUSTICE SYSTEM

Building on recent work in psychology, we have proposed an intuitive-override model of judging that is less idealistic than the "deductive" model embraced by the formalists, but also less cynical than the "intuitive" model embraced by the realists. The results of our CRT and judicial decision-making studies show that intuition influences judicial decision making. On balance, the judges we tested performed like other groups of well-educated adults—they largely based their judgments on intuition, but also demonstrated some ability to override intuition with deliberation.

Given the central role that judges play in the justice system both inside and outside the courtroom, reformers must understand judicial decision making before they can reshape the justice system to meet the needs of litigants and society. Our model raises two questions about judging. First, which decision-making approach—intuitive or deliberative—is more likely to produce accurate outcomes? Although we believe that intuition can be surprisingly accurate, we also believe that an excessive reliance on intuition will lead to erroneous judicial decisions. Second, how might the justice system encourage judges to make deliberative rather than intuitive decisions? We identify several concrete steps that might be taken.

A. Intuitive Versus Deliberative Decision Making

The intuitive approach to decision making is quick, effortless, and simple, while the deliberative approach to decision making is slow, effortful, and complex. The obvious advantage of the former is its speed; judges with heavy dockets can rely on intuition to make judgments quickly. The apparent advantage to the latter lies in the care it entails, suggesting that deliberative judgments are more likely to be accurate. Is this so?

Intuitive judgments are often quite accurate. As Daniel Kahneman and his long-time collaborator, Amos Tversky, observed in their early work on heuristics, intuitive thinking is "quite useful" and

can often lead to accurate decision making.[147] More recently, Gerd Gigerenzer and his colleagues have demonstrated that intuitive thinking can, with "a minimum of time, knowledge, and computation," enable decision makers "to make adaptive choices in real environments" in some circumstances.[148] And recent research suggests that some experts use intuitive thinking successfully. Consider, for example, the enormous body of research on chess grandmasters who routinely use intuitive rather than deliberative strategies to great effect:

> [T]he expert relies not so much on an intrinsically stronger power of analysis as on a store of structured knowledge. When confronted with a difficult position, a weaker player may calculate for half an hour, often looking many moves ahead, yet miss the right continuation, whereas a grandmaster sees the move immediately, without consciously analyzing anything at all.[149]

This conversion of deliberative judgment into intuitive judgment might be the hallmark of expertise.[150]

---

147 Tversky & Kahneman, *supra* note 99, at 1124; *see also* GLADWELL, *supra* note 1, at 14 ("[T]here are moments . . . when our snap judgments and first impressions can offer a much better means of making sense of the world.").

148 Gerd Gigerenzer & Peter M. Todd, *Fast and Frugal Heuristics: The Adaptive Toolbox*, *in* SIMPLE HEURISTICS THAT MAKE US SMART 3, 14 (Gerd Gigerenzer, Peter M. Todd & the ABC Research Group eds., 1999). For more on "fast and frugal" heuristics, see the other contributions in SIMPLE HEURISTICS THAT MAKE US SMART as well as in BOUNDED RATIONALITY: THE ADAPTIVE TOOLBOX (G. Gigerenzer & R. Selten eds., 2001). Some, like Gigerenzer, argue that fast-and-frugal heuristics yield results that are superior to those produced by deductive reasoning. *See* Epstein, *supra* note 23, at 719–21 (cataloguing various ways in which the "experiential" system can outperform the "rational" system).

149 Philip E. Ross, *The Expert Mind*, SCI. AM., Aug. 2006, at 64, 67; *see also* ADRIAAN D. DE GROOT, THOUGHT AND CHOICE IN CHESS (2d ed. 1978) (providing the classic study on decision making in chess); Kahneman & Frederick, *supra* note 28, at 51 ("Although System 1 is more primitive than System 2, it is not necessarily less capable. On the contrary, complex cognitive operations eventually migrate from System 2 to System 1 as proficiency and skill are acquired. A striking demonstration of the intelligence of System 1 is the ability of chess masters to perceive the strength or weakness of chess positions instantly."); Gary Klein, *The Fiction of Optimization*, *in* BOUNDED RATIONALITY: THE ADAPTIVE TOOLBOX, *supra* note 148, at 103, 115–16 (describing the "progressive deepening" strategy that expert chess players use in place of the traditional decision analysis); Tom Mueller, *Your Move*, THE NEW YORKER, Dec. 12, 2005, at 62, 64 ("Experienced players rely on subconscious faculties known variously as pattern recognition, visualization, and aesthetic sense. All are forms of educated guesswork—aids to making choices when certainty through exhaustive calculation is impossible—and may be summed up in a word: intuition. Even a novice player uses intuition to exclude most moves as pointless, and the more advanced a player becomes the less he needs to calculate."). More generally, research suggests that "[e]xperts notice features and meaningful patterns of information that are not noticed by novices." Nat'l Research Council, HOW PEOPLE LEARN: BRAIN, MIND, EXPERIENCE, AND SCHOOL 31 (John D. Bransford, Ann L. Brown & Rodney R. Cocking eds., 2000).

150 *See* HOGARTH, *supra* note 20, at 204 ("Many processes or reactions to stimuli that once relied heavily on the deliberate system can over time become automatic and thus bypass consciousness. This migration from the deliberate system to the tacit [system] is an important characteristic of the phenomenon of expertise.").

Nevertheless, there is reason to be suspicious of intuitive decision making in court. As Tversky and Kahneman observed, intuitive thinking also can "lead to severe and systematic errors."[151] In our own work, we have observed that judges who responded intuitively on the CRT made inaccurate choices.[152] On the judicial decision-making problems, which are obviously more relevant to appraising the relative merits of intuition versus deliberation in court, judges who employed intuitive thinking allowed an irrelevant settlement demand to influence their damage awards,[153] allowed an impressionistic assessment of statistical evidence to shape their liability determinations,[154] and allowed outcome information to influence their assessments of the ex ante predictability of appellate courts.[155] In these cases, intuitive, heuristic-based decision making led the judges to make erroneous decisions that they probably would have avoided had they adopted a deliberative approach.

Moreover, intuition is also the likely pathway by which undesirable influences, like the race, gender, or attractiveness of parties, affect the legal system.[156] Today, the overwhelming majority of judges in America explicitly reject the idea that these factors should influence litigants' treatment in court, but even the most egalitarian among us may harbor invidious mental associations.[157] For example, most white adults are more likely to associate African-Americans than white Americans with violence,[158] and most Americans are more likely to associate women with family life than with professional careers.[159] These associations seem to reflect automatic, intuitive judgments,[160] while active deliberation limits such biases.[161]

Furthermore, the capacity to use intuitive thinking successfully may require years of "effortful study"[162] as well as accurate and relia-

---

151 Tversky & Kahneman, *supra* note 99, at 1124.

152 *See supra* Part II.A.

153 *See supra* Part II.B.1.

154 *See supra* Part II.B.2.

155 *See supra* Part II.B.3.

156 *See* Christine Jolls & Cass R. Sunstein, *The Law of Implicit Bias*, 94 CAL. L. REV. 969, 976 (2006) (describing how race operates as a System 1 influence on judgment).

157 *See* Jerry Kang, *Trojan Horses of Race*, 118 HARV. L. REV. 1489, 1512–14 (2005) (reviewing the evidence on implicit invidious associations).

158 *See id.* at 1515 n.117.

159 *See* Jerry Kang & Mahzarin R. Banaji, *Fair Measures: A Behavioral Realist Revision of "Affirmative Action,"* 94 CAL. L. REV. 1063, 1072 (2006) ("[S]eventy-five percent of men and women do not associate female with career as easily as they associate female to family.").

160 *See* Jolls & Sunstein, *supra* note 156, at 973–74.

161 *See id.* at 974–75.

162 Ross, *supra* note 149, at 69 (describing K. Anders Ericsson's concept of "effortful study" whereby one "continually tackles challenges that lie just beyond one's competence"). *See generally* THE CAMBRIDGE HANDBOOK OF EXPERTISE AND EXPERT PERFORMANCE (K. Anders Ericsson, Neil Charness, Paul J. Feltovich & Robert R. Hoffman eds., 2006) (providing a collection of articles on the study of cognitive skill and expertise).

ble feedback on earlier judgments.[163] Unlike chess grandmasters, judges are unlikely to obtain accurate and reliable feedback on most of the judgments they make; indeed, they are only likely to receive external validation (or invalidation) of the accuracy of their judgments when their rulings are challenged on appeal.[164] The appeals process, however, does not provide reliable feedback. Many cases settle before appellate courts resolve the appeal; collateral policy concerns influence the outcome of some appeals, clouding the meaning of appellate decisions for the trial judge; and finally, appeals commonly take years to resolve, heavily diluting the value of any feedback. Moreover, the standards of review require appellate courts to give deference to trial judges on many of their discretionary decisions.[165] By the time an appellate court decides an appeal, the trial judge may have forgotten the nuances of the case, the law may have changed, or the judge may have retired or switched assignments. It is thus not surprising that we found no differences in CRT performance based on judges' experience or length of service. Unlike chess grandmasters, judges operate in an environment that does not allow them to perfect their intuitive decision-making processes.

Other aspects of the litigation process make it even more difficult for judges to receive good feedback. First, judges may have a narrow role in a case that precludes them from learning what happens later. For example, civil master calendar judges or criminal arraignment calendar judges might think that their decisions are correct, but they seldom learn how their decisions affect later proceedings in a case. Second, judges seldom receive useful feedback from lawyers or litigants. Lawyers usually say nothing to the judge about the quality of the judge's performance. When they do, judges easily discount such comments as biased. Third, judges are poorly positioned to learn how their decisions affect the world beyond the immediate case in front of them. Unlike legislators, they usually do not learn how their decisions affected subsequent conduct or events. Indeed, critics of the common-law process often base their criticism in part on this lack of feedback.[166]

Given the limitations of intuitive decision making, how exactly can judges (or anyone, for that matter) override their intuition with

---

163 *See* HOGARTH, *supra* note 20, at 208.

164 *See supra* note 17 and accompanying text. Overall, the rate of affirmance on appeal is quite high, at least in the federal courts. *See* Guthrie & George, *supra* note 17, at 361 fig.1b.

165 *See supra* note 18 and accompanying text.

166 *See* DONALD L. HOROWITZ, THE COURTS AND SOCIAL POLICY 53–56 (1977). *But see* Jeffrey J. Rachlinski, *Bottom-Up Versus Top-Down Lawmaking*, 73 U. CHI. L. REV. 933, 964 (2006) (noting that courts might be better capable of handling "social problems that produce vivid disasters" than are legislatures).

deliberation? Professor Robin Hogarth provides an excellent example:

> To illustrate, let us [ ]consider George, the dermatologist, who is examining a patient who has a growth below the right eye. When he first sees the growth, George has an immediate intuitive reaction. He has seen many growths in the past, although not necessarily below the right eye. However, the similarity between this growth and others of a particular type is striking. He just sees the resemblance without having to expend mental effort. This is George's tacit system in action. Yet George also knows that errors are made identifying growths. He therefore deliberately checks various features of this particular growth against a mental checklist in order to query his initial diagnosis. This second process is deliberative. It involves recalling details of codified medical knowledge. It involves attention and mental effort. This is the deliberative system at work.[167]

Consistent with this example, we do not suggest that judges should reject intuition in all cases. Rather, we suggest that judges should use deliberation as a verification mechanism especially in those cases where intuition is apt to be unreliable either because feedback is absent or because judges face cues likely to induce misleading reliance on heuristics.[168]

B. Inducing Deliberation

If judges need to use deliberation to override intuition, then the justice system should encourage that process. Of course, features of the existing justice system exist for many reasons, and efforts to encourage deliberation might have negative effects on other aspects of the system. Our goal here is simply to identify steps that the system could take to facilitate deliberation while recognizing that reformers would have to balance the benefits associated with these reforms against any costs they might impose.

Psychologists have determined that the decision-making approach one employs depends largely on "features of the task and of the individual."[169] Some features of a judge's tasks are not easily changed, but the environment in which judges perform their work could be more "kind," to use Robin Hogarth's term.[170] A "kind" environment allows intuition to flourish by providing immediate, high-quality feedback about the causes and consequences of errors.[171] Consider the following example of a "kind" environment:

---

167 HOGARTH, *supra* note 20, at 22.

168 *See* Guthrie et al., *supra* note 98, at 822–25.

169 Kahneman & Frederick, *supra* note 28, at 51.

170 *See* HOGARTH, *supra* note 20, at 89.

171 *See id.* at 88–89.

> The professional tennis player . . . [who] has played tennis almost every day since she was a child and has developed her intuitions about where to hit the ball on the basis of highly accurate feedback from coaches, personal experience in many matches (where the consequences of various shots are immediately and unmistakably apparent), and watching and analyzing other players' matches. Consequently, all her intuitions have been acquired within kind learning structures. There is no doubt in my mind that, during matches, her intuitions will be valid. (This does not mean, of course, that her choices will always be correct or that she will always execute shots precisely as she intends.)[172]

On the other hand, consider the following example of a "wicked" or "unkind" learning environment:

> Now consider physicians working in a hospital emergency room—a classic case of a wicked learning environment. Rarely do emergency room staff see the outcome of treatment. In cases of minor injury or illness, follow-up care is provided by the patient's own physician after discharge. In cases of more serious injury or illness, patients are transferred to a different ward for further evaluation and follow-up care. Emergency room staff therefore cannot rely on long-term feedback; they can observe only what happens in the short term. They also have few opportunities to experiment and often no way of knowing whether the task at hand is lenient (many conditions can be treated with antibiotics) or exacting (shortness of breath is symptomatic of many conditions, some minor, some potentially life threatening).[173]

Unfortunately, a judge's on-the-job intuitions generally develop in "wicked" environments. Judges sometimes function like emergency-room physicians in that they handle only part of a case. They may observe how well or how poorly things go while they are directly involved, but they often do not learn how things went at a later stage, so they cannot gauge the long-term effectiveness of their decisions. For example, a judge who decides whether to detain or release a particular defendant pending trial may not learn whether the defendant actually appeared for trial. In addition, errors seldom have direct adverse consequences for judges—when the judge slips, the litigant falls. This reality compounds the problems caused by the paucity of meaningful feedback. Even though most judges are conscientious and hard working, indirect consequences may be insufficient to guarantee good or improved performance. Reversal on appeal directly affects judges, but appeals occur infrequently and are seldom motivating.[174] Finally,

---

172 *Id.* at 89–90.

173 *Id.* at 218.

174 *See* David E. Klein & Robert J. Hume, *Fear of Reversal as an Explanation of Lower Court Compliance*, 37 LAW & SOC'Y REV. 579, 597–98 (2003); Donald R. Songer, Martha Hum-

although most judges want their colleagues to respect them,[175] one judge seldom learns the details of another judge's potentially erroneous decision making.

Little can be done about many of the aspects of the legal system that contribute to judges' "wicked" learning environments. The system can rectify other features, however, such as the amount of time that judges have to make rulings, the potential disciplining effect of opinion writing, the amount of training and feedback judges receive, the use of scripts, checklists, and multifactor tests in judging, and the allocation of decision-making resources.

1. *Time*

The justice system might expand the amount of time judges have to make decisions. Judges facing cognitive overload[176] due to heavy dockets[177] or other on-the-job constraints are more likely to make intuitive rather than deliberative decisions because the former are speedier and easier.[178] Furthermore, being cognitively "busy" induces judges to rely on intuitive judgment.[179] As many of the judges we have studied candidly admit, time pressures present an enormous challenge, often inducing less-than-optimal decision making.

No easy cure for time pressure exists, but the justice system could employ a few strategies to mitigate it. Most obviously, legislatures could expand the number of authorized judgeships in their jurisdictions, particularly in those areas with the heaviest dockets, thereby enabling judges to spend more time per case and per decision. This would be costly and it is unclear that the benefits would offset the costs. In particular, adding more judgeships might make litigation more attractive to those who would otherwise find alternative ways of resolving their disputes, just as adding more highways makes driving

---

phries Ginn & Tammy A. Sarver, *Do Judges Follow the Law when There Is No Fear of Reversal?*, 24 JUST. SYS. J. 137, 137 (2003).

175 *See* BAUM, *supra* note 11, at 50–60.

176 *See, e.g.*, BARRY SCHWARTZ, THE PARADOX OF CHOICE: WHY MORE IS LESS (2004) (exploring various ways in which the availability of many options and much information can compromise the quality of decision making); *see also* Chris Guthrie, *Panacea or Pandora's Box? The Costs of Options in Negotiation*, 88 IOWA L. REV. 601, 651 (2003) (concluding that multiple options can impose costs in decision making).

177 *See* Marc Galanter, *The Vanishing Trial: An Examination of Trials and Related Matters in Federal and State Courts*, 1 J. EMPIRICAL LEGAL STUD. 459, 461, 462–63 tbl. 1 (2004) (observing a five-fold increase in dispositions in federal courts from 1962–2002).

178 *See* Melissa L. Finucane, Ali Alhakami, Paul Slovic & Stephen M. Johnson, *The Affect Heuristic in Judgments of Risks and Benefits*, 13 J. BEHAV. DECISION MAKING 1, 8 (2000) (finding that subjects were more likely to rely on intuitive, heuristic-driven decision making rather than on deliberative decision making when operating under time pressure).

179 *See* Gilbert, *supra* note 46, at 179 ("The busyness-induced undercorrection of dispositional inferences is now a well-established and widely replicated phenomenon.").

more attractive to those who would otherwise find alternative means of transportation.[180]

Minimizing the number of spur-of-the-moment decisions that judges are expected to make might also help. Our model suggests that decisions made during pretrial conferences, settlement conferences, motion hearings, and so forth are more likely to be intuitive and impressionistic rather than deliberative and well reasoned. Likewise, evidentiary rulings at trial are apt to be more prone to error than are those rulings made before trial. When ruling on the admissibility of evidence at trial, judges often have little choice but to think intuitively. Our model suggests that judges should not make important evidentiary rulings in such a setting. To be sure, motions *in limine* deprive the judge of the context in which the evidence will be heard. Nonetheless, we suggest that judges might require parties to file their more important evidentiary motions before trial, but delay ruling on them until the issues arise during the trial, and even then after a recess in which the judge has had some time to study the papers and deliberate.

Taking time to deliberate rather than relying on intuition might not always produce better judgments. For example, if a judge has to determine whether a witness is being truthful, the judge's intuitive decision based on observing the witness's behavior might be more accurate than a deliberative decision made in chambers days after the details have faded. On the other hand, there are times when careful deliberation is desirable. If judges are susceptible to the "beauty bias," for example, they might evaluate an attractive witness's credibility too positively and an unattractive witness's credibility too negatively if they make a hasty judgment in the courtroom.[181] A reflective determination made in chambers after the impact of the witness's appearance has worn off might be more accurate.

2. *Opinion Writing*

The justice system also might require judges to write opinions more often.[182] In some respects, this prescription conflicts with the previous measure because opinion writing takes an enormous amount of time, which judges might not have. Despite this cost, writing opin-

---

180 *See* Tracey E. George & Chris Guthrie, *Induced Litigation*, 98 NW. U. L. REV. 545, 547 (2004).

181 *See* Judith H. Langlois, Lisa Kalakanis, Adam J. Rubenstein, Andrea Larson, Monica Hallam & Monica Smoot, *Maxims or Myths of Beauty? A Meta-Analytic and Theoretical Review*, 126 PSYCHOL. BULL. 390, 399–401 (2000).

182 Judges generally disclose the reasons behind their actions, and they are often required to do so. *See, e.g.*, FED. R. CIV. P. 52(a) (requiring findings of fact after a bench trial). Arguably, judges already explain the reasons for their decisions more frequently and completely than any other public officials.

ions could induce deliberation that otherwise would not occur. Rather than serving merely to describe an allegedly deliberative process that has already occurred (as the formalists might argue) or to rationalize an intuitive decision already made (as the realists might argue), the discipline of opinion writing might enable well-meaning judges to overcome their intuitive, impressionistic reactions.[183] The process of writing might challenge the judge to assess a decision more carefully, logically, and deductively.[184]

Preparing a written opinion is sometimes too inconvenient or simply infeasible. In such situations, perhaps judges should be required to articulate the basis for the decision before announcing the conclusion. Though there is little opportunity for reflection in these situations, simply stating the reasons for the decision before the ruling is announced may encourage the judge to be more deliberative.

The psychological literature on the effects of requiring decision makers to provide reasons is mixed. Providing reasons for one's decision induces deliberation,[185] but can also impair performance on tasks that benefit from intuition.[186] Some studies suggest that deliber-

---

183 Some have encouraged the preparation of written opinions for exactly this reason. *See, e.g.*, Moses Lasky, *Observing Appellate Opinions from Below the Bench*, 49 CAL. L. REV. 831, 838 (1961) ("Where a judge need write no opinion, his judgment may be faulty. Forced to reason his way step by step and set down these steps in black and white, he is compelled to put salt on the tail of his reasoning to keep it from fluttering away."); Robert A. Leflar, *Some Observations Concerning Judicial Opinions*, 61 COLUM. L. REV. 810, 810 (1961) ("[T]he necessity for preparing a formal opinion assures some measure of thoughtful review of the facts in a case and of the law's bearing upon them. Snap judgments and lazy preferences for armchair theorizing . . . are somewhat minimized."); Roger J. Traynor, *Some Open Questions on the Work of State Appellate Courts*, 24 U. CHI. L. REV. 211, 218 (1957) ("In sixteen years I have not found a better test for the solution of a case than its articulation in writing, which is thinking at its hardest. A judge, inevitably preoccupied with the far-reaching effect of an immediate solution as a precedent, often discovers that his tentative views will not jell in the writing. He wrestles with the devil more than once to set forth a sound opinion that will be sufficient unto more than the day."). *But see* Chad M. Oldfather, *Writing, Cognition, and the Nature of the Judicial Function*, 96 GEO. L.J. (forthcoming Apr. 2008) (arguing generally that verbalization does not always enhance understanding or decision making, particularly when important aspects of the situation are not readily susceptible to verbalization).

184 *See* HOGARTH, *supra* note 20, at 262 ("[W]hen explicitly asked about the rationale behind their choices, people will change their choices if those choices are not consistent with the reasons that they are able to produce. Thus, if deliberation were structured to highlight 'good' reasons, outcomes might well be better than initial intuition.").

185 *See id.* at 263 ("Verbalization . . . forces people to act in [a] deliberate mode and cuts off access to tacit processes.").

186 *See* John McMackin & Paul Slovic, *When Does Explicit Justification Impair Decision Making?*, 14 APPLIED COGNITIVE PSYCHOL. 527, 535–39 (2000) (finding that asking subjects to provide reasons adversely affected their performance on intuitive tasks, such as indicating which advertisement people would prefer, but improved their performance on analytical tasks, such as estimating the length of the Amazon River); Timothy D. Wilson & Jonathan W. Schooler, *Thinking Too Much: Introspection Can Reduce the Quality of Preferences and Decisions*, 60 J. PERSONALITY & SOC. PYSCHOL. 181, 181 (1991) (finding that subjects' preferred choices of strawberry jam were less likely to correspond with experts' preferred choices if the subjects were required to give reasons for their choices).

ation can sometimes produce results that are inferior to results produced by intuition, particularly where a task involves aesthetic judgment.[187] We suspect, however, that most of the judgments that judges make are not the sort that are impaired by deliberation.

3. *Training and Feedback*

Just as there is continuing legal education for lawyers, there also is continuing legal education for judges. The justice system could provide even more training opportunities for judges and invest more resources in the kinds of judicial training most likely to facilitate deliberative decision making. Several studies suggest that statistical training in particular, though certainly not a panacea,[188] increases the likelihood that individuals will make rational, deliberative decisions rather than intuitive, heuristic-driven ones.[189]

Training could help judges understand the extent of their reliance on intuition and identify when such reliance is risky—the necessary first steps in self-correction. Judges could learn to interrupt their intuition, thereby allowing deliberation to intervene and modify behavior, if not actually altering underlying prejudices or attitudes.[190] Additionally, providing judges with instruction and practice in scientific reasoning hopefully would teach them to employ deductive reasoning either deliberatively or, if the instruction and practice is

---

187 *See* Timothy D. Wilson, Douglas J. Lisle, Jonathan W. Schooler, Sara D. Hodges, Kristen J. Klaaren & Suzanne J. LaFleur, *Introspecting About Reasons Can Reduce Post-Choice Satisfaction*, 19 PERSONALITY & SOC. PSYCHOL. BULL. 331, 337 (1993) (finding that experimental subjects who chose a poster to take home based on intuition were more pleased with their choice than those who were required to deliberate before selecting a poster).

188 *See, e.g.*, *Small Numbers*, *supra* note 114, at 107–09 (finding evidence of intuitive and heuristic-driven thinking even among mathematically oriented psychologists).

189 *See, e.g.*, GERD GIGERENZER, CALCULATED RISKS: HOW TO KNOW WHEN NUMBERS DECEIVE YOU 229–46 (2002) (exploring how the population can be taught to understand risk information); Franca Agnoli, *Development of Judgmental Heuristics and Logical Reasoning: Training Counteracts the Representativeness Heuristic*, 6 COGNITIVE DEV. 195, 195 (1991) (showing that statistical training improved children's reasoning abilities); Franca Agnoli & David H. Krantz, *Suppressing Natural Heuristics by Formal Instruction: The Case of the Conjunction Fallacy*, 21 COGNITIVE PSYCHOL. 515, 515 (1989) (showing that statistical training improved the reasoning abilities of adult subjects with limited prior exposure to mathematics); Richard P. Larrick, James N. Morgan & Richard E. Nisbett, *Teaching the Use of Cost-Benefit Reasoning in Everyday Life*, 1 PSYCHOL. SCI. 362, 362 (1990) (showing that cost-benefit training improved decision making); Richard E. Nisbett, David H. Krantz, Christopher Jepson & Ziva Kunda, *The Use of Statistical Heuristics in Everyday Inductive Reasoning*, 90 PSYCHOL. REV. 339, 339 (1983) (identifying several ways in which statistical training can improve the intuitive reasoning people generally employ); *see also* HOGARTH, *supra* note 20, at 23 ("[I]ntuition is like expertise. It is specific to particular domains. It is acquired through domain-relevant experience. And it can be improved through instruction and practice.").

190 *See* HOGARTH, *supra* note 20, at 209 ("Just as we cannot avoid tacitly forming prejudices, we cannot avoid forming a good first impression of con men. But we can learn not to act uncritically on the basis of that first impression.").

especially effective, automatically.[191] Judges also could receive help identifying which aspects of their environments are "wicked" and how to improve them.

Likewise, jurisdictions could adopt peer-review processes to provide judges with feedback. For example, every two years, three experienced judges from other jurisdictions could visit a target court. They could select a few cases recently decided by each target court judge, read all of the rulings and transcripts, and then provide the judges with feedback on their performance and constructive suggestions for improvement. This would give judges an opportunity to obtain feedback on issues that typically escape appellate review. When aggregated, the results of such a process might also identify structural problems that amendments to rules or statutes should remedy. Such a procedure also would increase judicial accountability by subjecting decisions that escape appellate review to peer review. Research has shown that accountability of this sort can improve decision-making performance.[192] If a peer-review process is infeasible, courts could at a minimum record and provide judges with outcome data on relevant decisions—for example, whether a defendant released on bail actually appeared for trial. Armed with this feedback, judges might be better able to learn what they are doing well and what they are doing poorly.

Of course, most judges are generalists,[193] which might impede their efforts to learn good decision-making skills and to accept meaningful feedback through training and peer-review processes. With the exception of the tasks judges perform repeatedly, it might take a long time for judges to acquire sufficient experience in handling a particular issue to accumulate enough feedback to avoid errors. It is as if a professional tennis player divided his or her time among tennis, vol-

---

191 *See* Evan R. Seamone, *Judicial Mindfulness*, 70 U. CIN. L. REV. 1023, 1074–75 (2002) (advocating that judges engage in an elaborate program of self-examination and deliberation in an effort to eliminate or mitigate bias resulting from automatic, mindless impulses). The author's program strikes us as somewhat impractical, even though it is motivated by concerns very much like our own.

192 *See, e.g.*, Jennifer S. Lerner & Philip E. Tetlock, *Accounting for the Effects of Accountability*, 125 PSYCHOL. BULL. 255, 256–59 (1999). The authors explain that decision makers are more likely to engage in self-critical thinking if they learn prior to making their decisions that they will be accountable to an audience whose views are unknown, who is well-informed, and who has a legitimate reason for evaluating the decision makers' judgments. *See id.* at 259.

193 *See* David L. Faigman, *Judges as "Amateur Scientists,"* 86 B.U. L. REV. 1207, 1209 (2006) ("By its nature, law requires judges to be generalists"); William K. Ford, *Judging Expertise in Copyright Law*, 14 J. INTELL. PROP. L. 1, 3 (2006) ("[J]udges are generalists who have jurisdiction over an enormous range of legal disputes: copyright law one day, environmental law the next, antitrust the day after that."); Rachlinski et al., *supra* note 98, at 1228 ("Judges . . . have historically been generalists who preside over any and all cases."); Jonathan Fabian Witt, *Bureaucratic Legalism, American Style: Private Bureaucratic Legalism and the Governance of the Tort System*, 56 DEPAUL L. REV. 261, 265 (2007) (describing "reliance on generalist judges" as one of the "[c]entral features of the common-law system").

leyball, softball, soccer, and golf rather than concentrating on tennis—the player's opportunity to develop "tennis intuition" would diminish. Although we have concluded elsewhere that specialization may not insulate judges from cognitive illusions such as anchoring,[194] it might mitigate such biases by maximizing the opportunity to benefit from a large quantity of relevant feedback. Moreover, because the benefit of experiential learning in a wicked environment is limited, training may be necessary to compensate for deficiencies in the learning environment.[195]

4. *Scripts, Checklists, and Multifactor Tests*

Scripts and checklists can free judges from reliance on their memories and encourage them to proceed methodically, thereby ensuring that they touch all of the deliberative bases. A judge who must review a script or checklist at each step in the decision-making process is less likely to rely on intuition when doing so is inadvisable.

In some respects, the justice system already takes this approach. Judges receive "scripts" for some tasks when they are appointed.[196] Judges also develop their own scripts and checklists for various tasks and share them with one another.[197] In addition, multifactor or balancing tests have become common. Examples include the five-factor test for inadvertent waiver of privileged documents;[198] the seven- or eight-factor test for cost shifting in the context of electronic discovery;[199] the five-factor test for dismissal for failure to prosecute or comply with an order;[200] the four-factor test for evaluating speedy trial right violations;[201] and the multifactor test for determining the likelihood of consumer confusion in trademark infringement cases.[202]

---

194 *See* Rachlinski et al., *supra* note 98, at 1236–37. *But see* Chris Guthrie & Jeffrey J. Rachlinski, *Insurers, Illusions of Judgment & Litigation*, 59 VAND. L. REV. 2017, 2022 (2006) (finding that insurers are less susceptible than lay litigants to the anchoring effect).

195 Baruch Fischhoff, *Heuristics and Biases in Application*, *in* HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT, *supra* note 23, at 730, 731 ("Training provides feedback that everyday life typically lacks, allowing people to test and refine judgment skills.").

196 *See, e.g.*, FED. JUDICIAL CTR., BENCHBOOK FOR U.S. DISTRICT COURT JUDGES (4th ed. 1996, rev. 2000).

197 *See* E-mail correspondence among Magistrate Judges of the United States District Court for the Central District of California (Dec. 15, 2004) (on file with author).

198 *See, e.g.*, United States *ex rel.* Bagley v. TRW, Inc., 204 F.R.D. 170, 177–78 (C.D. Cal. 2001).

199 *See, e.g.*, Rowe Entm't, Inc. v. William Morris Agency, Inc., 205 F.R.D. 421, 428–29 (S.D.N.Y. 2002).

200 *See, e.g.*, Valley Eng'rs, Inc. v. Elec. Eng'g Co., 158 F.3d 1051, 1056–57 (9th Cir. 1998).

201 *See, e.g.*, Barker v. Wingo, 407 U.S. 514, 530 (1972).

202 *See, e.g.*, Jada Toys, Inc. v. Mattel, Inc., 496 F.3d 974, 979 (9th Cir. 2007) (applying the multifactor test developed in AMF Inc. v. Sleekcraft Boats, 599 F.2d 341, 348–49 (9th Cir. 1979), and criticizing the practice of considering merely some factors).

Although multifactor tests are ubiquitous, they are imperfect. Some multifactor tests are poorly designed and include inappropriate factors that duplicate or overlap with other factors within the test. When judges excessively rely on multifactor tests, as well as on scripts and checklists, there is a risk of mechanical jurisprudence. Excessive rigidity may unduly restrict judges from tailoring their analysis to the case. Further, multifactor or balancing tests may be indeterminate, and applying or weighing some of the factors within the test may require intuition. Finally, judges sometimes employ heuristics to circumvent the multifactor analysis by relying on just a few of the factors in making their decision, thereby diminishing the value of the test as a corrective device.[203]

Nevertheless, such tests possess the potential for mitigating cognitive error by nudging judges toward more deliberative processes. Multifactor tests can help ensure that judges consider all relevant factors and can remind them of their responsibility to base decisions on more than mere intuition.[204] Similar reminder systems have reduced medical diagnostic error.[205] Thus, a system that forces judges to weigh each of the factors expressly also might help reduce judges' reliance on intuition. This may be the reason that some appellate courts require administrative agencies or lower courts to expressly consider or weigh each of the factors in a multifactor test, sometimes in a particular sequence.[206] In their more extreme forms, such techniques are

---

203 *See* Barton Beebe, *An Empirical Study of the Multifactor Tests for Trademark Infringement*, 94 CAL. L. REV. 1581, 1581–82 (2006).

204 *See* Chip Heath, Richard P. Larrick & Joshua Klayman, *Cognitive Repairs: How Organizational Practices Can Compensate for Individual Shortcomings*, 20 RES. ORGANIZATIONAL BEHAV. 1, 15 (1998) ("[I]ndividuals attend to and process information more comprehensively when they have a mental *schema* that tells them what information is needed in a give situation and where to find it.").

205 *See* Padmanabhan Ramnarayan, Andrew Winrow, Michael Coren, Vasanta Nanduri, Roger Buchdahl, Benjamin Jacobs, Helen Fisher, Paul M. Taylor, Jeremy C. Wyatt & Joseph Britto, *Diagnostic Omission Errors in Acute Paediatric Practice: Impact of a Reminder System on Decision-Making*, 6 BMC MED. INFORMATICS & DECISION MAKING 37, 37–38 (2006), *available at* http://www.biomedcentral.com/1472-6947/6/37 (reporting that physicians' use of an Internet-based diagnostic reminder system improved diagnostic workups and reduced diagnostic omission errors; the reminder system reduced unsafe diagnostic workups from 45.2% to 32.7%); *see also* Paul R. Dexter, Susan Perkins, J. Marc Overhage, Kati Maharry, Richard B. Kohler & Clement J. McDonald, *A Computerized Reminder System to Increase the Use of Prevention Care for Hospitalized Patients*, 345 NEW ENG. J. MED. 965, 965 (2001) (reporting similarly positive results from use of a computerized reminder system to remind physicians to deliver preventive care to hospital patients).

206 *See, e.g.*, Ng v. I.N.S., 804 F.2d 534, 538 (9th Cir. 1986) ("[W]e require that the BIA state its reasons and show proper consideration of all factors when weighing equities . . . . [T]his court cannot assume that the BIA considered factors that it failed to mention in its decision."); Educ. Credit Mgmt. Corp. v. Pope, 308 B.R. 55, 59 (N.D. Cal. 2004) ("[T]he Ninth Circuit adopted a three-part undue hardship test . . . . [C]ourts must consider each element [of the test] in turn and, where one of the three elements is not met, the court must stop there with a finding of no dischargability."); Frankel v. Frankel, 886 A.2d 136,

known as "forcing functions," which are exemplified by computer systems that force the user to complete step two before moving to step three.[207]

5. *Reallocation of Decision-Making Authority*

The justice system might also take procedural steps to limit judges' exposure to stimuli that are likely to trigger intuitive thinking. This might be done in several ways, but we think the most promising approach is for courts to expand their use of what we call "divided decision making." By way of illustration, consider the first anchoring study described above, in which a preposterous $10 million demand made at a settlement conference influenced judges' damage awards.[208] By dividing decision making between judges—for example, by assigning managerial functions to one judge and adjudicative functions to another—the justice system could prevent such an anchor from influencing a trial judge's determination of damages. The justice system could also achieve this end by separating decision-making authority between judge and jury—for example, by exposing the judge, but not the jurors, to potentially misleading inadmissible evidence. More generally, this divided decision-making strategy is likely to shield the ultimate adjudicator from various stimuli that are likely to induce intuitive, heuristic-based decision making.

There are some impediments to expanding the use of divided decision making beyond existing boundaries. For example, litigants are entitled to jury trials in only certain cases. Perhaps more significantly, the costs of some forms of divided decision making, such as assigning two judges to every case, could be quite high. Still, this approach has much to offer.

6. *Summary*

In short, there are numerous ways in which the civil and criminal justice systems can increase the likelihood that judges will make more deliberative decisions. Each of these reforms tends to make decision making more costly or time consuming, and some of the reforms,

---

154 (Md. App. 2005) ("A trial judge must consider each factor listed . . . when determining the amount of monetary award.").

207 *See* Mads Soegaard, Forcing Functions, http://www.interaction-design.org/encyclopedia/forcing_functions.html (last visited Sept. 25, 2007) ("A forcing function is an aspect of a design that prevents the user from taking an action without *consciously* considering information relevant to that action. It *forces* conscious attention upon something . . . and thus deliberately disrupts the efficient or automatised [sic] performance of a task. . . . It is . . . useful in situations where the behavior of the user is *skilled*, as in performing routine or well-known tasks. Execution of this type of task[ ] is often partly or wholly automatized, requiring few or no attentional resources . . . , and it can thus be necessary to 'wake the user up' by deliberately disrupting the performance of the task.").

208 *See supra* Part II.B.1.

such as divided decision making, might be sufficiently cumbersome that they do not justify the extra costs imposed on litigants and the justice system. Still, gains in accuracy, and therefore justice, may be worth the costs of reform.

CONCLUSION

We believe that most judges attempt to "reach their decisions utilizing facts, evidence, and highly constrained legal criteria, while putting aside personal biases, attitudes, emotions, and other individuating factors."[209] Despite their best efforts, however, judges, like everyone else, have two cognitive systems for making judgments—the intuitive and the deliberative—and the intuitive system appears to have a powerful effect on judges' decision making. The intuitive approach might work well in some cases, but it can lead to erroneous and unjust outcomes in others. The justice system should take what steps it can to increase the likelihood that judges will decide cases in a predominately deliberative, rather than a predominately intuitive, way.

In his recent book, *How Doctors Think*, Dr. Jerome Groopman observes that "[m]uch has been made of the power of intuition, and certainly initial impressions formed in a flash can be correct. But as we hear from a range of physicians, relying too heavily on intuition has its perils. Cogent medical judgments meld first impressions—gestalt—with deliberate analysis."[210] Like cogent medical judgments, cogent legal judgments call for deliberation. Justice depends on it.

---

209 Donald C. Nugent, *Judicial Bias*, 42 CLEV. ST. L. REV. 1, 4 (1994). For a less optimistic take on judicial performance, see Geoffrey P. Miller, *Bad Judges*, 83 TEX. L. REV. 431, 431 (2004) (examining "bad judges" who are "incompetent, self-indulgent, abusive, or corrupt").

210 JEROME GROOPMAN, HOW DOCTORS THINK 9 (2007).