

Annual Review of Law and Social Science

Judging the Judiciary by the Numbers: Empirical Research on Judges

Jeffrey J. Rachlinski¹ and Andrew J. Wistrich²

¹Henry Allen Mark Professor of Law, Cornell Law School, Ithaca, New York 14853

²United States District Court, Central District of California, Los Angeles, California 90012

Annu. Rev. Law Soc. Sci. 2017. 13:203–29

The *Annual Review of Law and Social Science* is online at lawsocsci.annualreviews.org

<https://doi.org/10.1146/annurev-lawsocsci-110615-085032>

Copyright © 2017 by Annual Reviews.
All rights reserved

Keywords

judges, courts, judicial decision making

Abstract

Do judges make decisions that are truly impartial? A wide range of experimental and field studies reveal that several extralegal factors influence judicial decision making. Demographic characteristics of judges and litigants affect judges' decisions. Judges also rely heavily on intuitive reasoning in deciding cases, making them vulnerable to the use of mental shortcuts that can lead to mistakes. Furthermore, judges sometimes rely on facts outside the record and rule more favorably toward litigants who are more sympathetic or with whom they share demographic characteristics. On the whole, judges are excellent decision makers and sometimes resist common errors of judgment that influence ordinary adults. The weight of the evidence, however, suggests that judges are vulnerable to systematic deviations from the ideal of judicial impartiality.



ANNUAL REVIEWS Further

Click [here](#) to view this article's online features:

- Download figures as PPT slides
- Navigate linked references
- Download citations
- Explore related articles
- Search keywords

INTRODUCTION

Judges are the axle on which the wheels of justice turn. They manage pretrial proceedings, mediate settlement conferences, rule on motions, conduct bench trials, supervise jury trials, take guilty pleas, impose criminal sentences, and resolve appeals. In the process, they find facts, make or apply law, and exercise discretion. Judges wield enormous power, and society therefore rightly expects much of them. Judges must be fair-minded, impartial, patient, wise, efficient, and intelligent (Wistrich 2010). They must set aside their politics and their prejudices, make rational decisions, and follow the law (see, e.g., Am. Bar Assoc. 2011, rules 1.1, 1.2, 2.2–2.5, 2.8). But is it possible for judges to perform as we expect?

The answer to this question remains somewhat uncertain. Twenty years ago, Lawrence Baum (1997, p. 149) concluded, “Despite all the progress that scholars have made, progress that is accelerating today, we are a long way from achieving truly satisfying explanations of judicial behavior.” Much more research has been conducted since then, but judicial behavior still remains something of a mystery. Some scholars argue that judges behave rationally but make decisions that further their self-interest (Epstein et al. 2013). That assertion, however, raises as many questions as it answers: What do judges see as their self-interest? Are fairness and impartiality their primary goals? What incentives do judges really face? After all, they rarely lose their positions and seldom get promoted. And even if judges primarily strive for fairness and impartiality, do they achieve these goals?

Research on human judgment and choice indicates that most people face cognitive limitations that lead them to make choices that do not consistently further their own ends (Ariely 2009). People commonly rely on intuition and simple shortcuts (or heuristics) to make choices (Kahneman 2011). Heuristics can be effective and surprisingly accurate (Gigerenzer & Todd 1999) but can also lead to predictable mistakes when over-applied or misused. These problems plague professionals as well. Research on doctors, dentists, accountants, futures traders, and others shows that they all fail to live up to an idealized standard of judgment in many settings (Ariely 2009). It would be surprising if judges are any different.

The available research on judges suggests that they sometimes fall short of the lofty ideal to which society holds them. A growing body of research supports the conclusion that although judges are often excellent decision makers, they have vulnerabilities. At the outset, we know that in some areas of law, judicial decisions are too chaotic. A study of immigration asylum decisions, for example, reveals that some judges grant asylum in a high percentage of cases, whereas others almost never grant asylum (Ramji-Nogales et al. 2007). Asylum outcomes thus turn on the random assignment of a case to one judge or another. Decisions concerning whether to grant leave to appeal or to allow release on bond in immigration cases are similarly erratic (Rehaag 2012, Ryo 2016). Concerns about variation in conviction rates have also long haunted criminal law (Weisselberg & Dunworth 1993). Even in criminal sentencing decisions in federal court, in which a highly structured set of guidelines constrains judges, variation remains robust (Scott 2011). Judges do not seem to decide as reliably as might be hoped or expected. Worse still, the variation does not just arise from chaos or a lack of meaningful standards; it arises from systematic vulnerabilities in how judges think.

This article surveys the empirical research that assesses whether judges live up to the standards of their profession. The evidence accumulated to date reveals that judges fall short in predictable ways. First, as the legal realists feared, judges’ personal characteristics influence their decision making. Specifically, the research indicates that when cases raise issues that are salient to judges’ personal characteristics, they do not consistently put their characteristics aside. Second, judges overreact to mechanisms of accountability, such as appellate review, retention, and promotion.

Third, judges rely too heavily on intuitive ways of thinking that can be misleading. Fourth, in making decisions, judges sometimes rely on factors outside the record, including inadmissible evidence, their emotional reactions, and prejudices.

To be fair to judges, they labor under a great deal of academic scrutiny. The existing research on judicial decision making probably focuses too heavily on judicial failings. Scholars conduct their research with an eye toward showing that judges are politically motivated or biased. This is understandable, given the ideal of neutral judging that society expects from judges, but the emphasis on deviations likely makes judges seem worse than they are. The research includes several studies in which judges adhere to an ideal norm of neutrality, and we certainly include these in our review. No studies really provide usable estimates of how many cases are skewed by politics, prejudice, or other misjudgment, and the research does not support a means of making a reasonable estimate. The circumstances under which judges deviate from the norm are nevertheless worth exploring, not to make judges look bad, but to identify potential ways they might improve.

In reaching our conclusions, we review a diverse array of both experimental and field studies of judicial decision making. We set aside judges' autobiographies and biographies, interviews of judges, careful parsing of individual opinions, and judges' own accounts of how they make decisions. Such undertakings can provide valuable insights, but our focus lies on systematic empirical accounts of judicial decision making. These include archival studies of actual decisions and experiments or simulations using hypothetical cases. Although most research on judges emphasizes decisions of the US Supreme Court (especially since the Second World War), our focus lies with the state courts, lower federal courts, and a handful of foreign courts. Although the US Supreme Court is important, of course, it resolves few cases and represents only a tiny window into the judicial decision-making process. Each of the studies we incorporate into our analysis involves vastly more judges than the 39 people who have served on the Supreme Court in the last 70 years. The focus on the Supreme Court also tends to emphasize the role of politics in judging. Political influence is only one way judges can fail to meet the demands of their roles. We discuss this concern but expand upon it.

THE INFLUENCE OF JUDICIAL PERSONALITY

It does not take a careful social scientific inquiry to see that judges do not always agree on how cases should be decided. To be sure, many of the cases that judges decide are straightforward, and unanimous opinions in the appellate courts are the norm (Posner 2008). Nevertheless, the nature of the litigation process ensures that judges also must decide many close cases (Priest & Klein 1984). Factual accounts conflict, and reasonable judges might disagree on which account is more accurate. Also, the law leaves a great deal of space for interpretation. Those who appear before judges, however, are entitled to reliable and impartial rulings. Nevertheless, individual variation arising from personal characteristics of judges influences how judges decide cases (Sisk et al. 1998). As the research demonstrates, personal characteristics of judges—their political ideology, gender, race, and experience—affect their decisions in cases that reflect those characteristics.

Political Ideology

Political ideology is the most widely cited extralegal influence on judges (for reviews, see Epstein et al. 2013, Maveety 2003, Pinello 1999). An avalanche of research demonstrates that US Supreme Court Justices make decisions that align with their political attitudes (Epstein & Knight 1997, Segel & Spaeth 2015). Lower federal court judges display similar tendencies (Sunstein et al. 2006). A

thorough review of this literature lies beyond the scope of our inquiry here, and we focus on other aspects of judicial decision making. The literature on judicial political ideology has been reviewed at length elsewhere, and even the leading exponents of political ideology now acknowledge that focusing too narrowly on it provides an incomplete account of judicial behavior (Epstein 2016, Epstein & Knight 2013). We nevertheless accept the conclusion that judges are political actors whose beliefs influence their judgment. That said, we note that most of the research here focuses on the Supreme Court and the highest state courts. Studies of trial court judges are less common, and notably, some show a more mixed portrait of political influence in the trial courts (Blume & Eisenberg 1999). Inasmuch as more than 80% of cases are ultimately decided by the trial court alone (Zorn & Bowie 2010), this result gives some comfort that for most litigants their cases will not be decided largely by the politics of the presiding judge.

If judges are political actors, then their political attitudes are apt to be most salient in cases with direct political implications. Cox & Miles (2008) found exactly that result in voting rights cases. They concluded that the political party of the presiding judge has a large influence on case outcomes among federal judges. These authors reported that Democratic judges are more likely to find liability under Section 2 of the Voting Rights Act (VRA) than their Republican counterparts. Section 2 cases involve claims that redistricting patterns have undermined minority voting rights, which would disfavor Democrats. Cox & Miles (2008) found that African American judges are also especially receptive to such claims, being twice as likely as other judges to find a violation of the VRA. Furthermore, the presence of one Democrat or one African American on the panel influences the voting patterns of the other two judges—a pattern that researchers have identified in several other settings (Farhang & Wawro 2004). Political attitudes thus seem to influence judges when politics lies at the heart of a case.

Demographics and Judging

The legal realists argued that a judge's personal characteristics influence how they decide cases (Frank 1930). Contemporary research supports and sharpens their point. As we describe below, demographic characteristics—e.g., religion, gender, race, and past employment—all influence judges. The influence of demographics follows a predictable pattern: Judges tend to decide cases that present issues salient to their demographic characteristics in ways that favor their demographic profile.

Religion. Judges' religious backgrounds influence their decision making (Idelman 2005). Jewish and Catholic judges display more politically liberal trends in decision making than their Evangelical and Protestant counterparts (see Bornstein & Miller 2009 for a review), but this finding does not discern whether religion produces different general political attitudes or whether other demographic factors that correlate with religion are at play. Other studies demonstrate that specific religious values lead judges to decide certain kinds of cases differently. For example, consistent with papal teachings, Catholic judges (along with Evangelical judges) are more likely to side against gay rights relative to other judges (Pinello 2003) but are more moderate in capital cases (Songer & Tabrizi 1999). Catholics and Evangelicals likewise have been shown to be harsher on defendants in obscenity cases (Songer & Tabrizi 1999). Religious orientation also influences judges heavily in religious accommodation cases: "Jewish judges vote heavily separationist, Catholics vote heavily accommodationist, and Protestants divide" (Sorauf 1976, at p. 220). A more recent study confirmed the result that Jewish judges and other judges who follow less common faiths much more strongly favor accommodating minority religious practices than Catholic and Protestant judges

(Sisk et al. 2004). In effect, religious orientation matters when core aspects of the judges' religion are at play in the cases they decide.

Gender. After women began entering the ranks of the judiciary in appreciable numbers, scholars began asking whether their gender would influence their decisions on the bench. The results are equivocal. Two early studies of sentencing in urban trial courts found no differences between male and female judges (Gruhl et al. 1981, Kritzer & Uhlman 1977). More recent work replicated this finding (Steffensmeier & Hebert 1999). Other studies also found little effect of gender among federal appellate judges (Davis 1986, Gottschall 1983). Overall, female judges are not particularly more or less conviction prone than their male counterparts, nor do they clearly favor or disfavor plaintiffs in civil cases.

Gender matters in cases in which gender itself is an issue, however. Songer et al. (1994), for example, found that the gender of federal appellate judges did not influence decisions in obscenity or search-and-seizure cases, but female judges favored plaintiffs in employment discrimination cases relative to their male counterparts. Differences are also more likely to be found in cases involving so-called women's issues or in which gender is salient—mostly workplace discrimination and harassment cases (Allen & Wall 1987, Boyd 2013, Boyd et al. 2010, Farhang & Wawro 2004). In one notable study, the presence of a single female judge on a three-judge panel in the US Courts of Appeals increased the success rate for female plaintiffs in sexual discrimination claims relative to all-male panels (Perisie 2005). The result is striking, as a single judge can easily be outvoted by the other two. The finding suggests that the presence of a female judge adds a different perspective that disproportionately alters the ultimate outcome. Even in cases involving gender issues, however, studies do not uniformly report differences between male and female judges (Walker & Barrow 1985).

A recent related study expands the notion that demographic characteristics like gender can influence judgment at a personal level. The study found that in gender-related cases (such as sexual harassment), a federal judge with one or more daughters is 7% more likely to vote in favor of the female plaintiff compared with a judge with no daughters (Glynn & Sen 2015). In cases not involving gender-related issues, however, judges with daughters do not differ from their counterparts who have either no children or only male children. Apparently, judges come to identify with their daughters' interests.

Race. As with gender, a judge's race seems to matter most when race is a central issue in the case. In the study by Cox & Miles (2008) that we discussed above, the authors found that the effect of race on the outcome of Section 2 claims under the VRA was much stronger than the effect of politics alone. Similarly, African American federal district judges more often rule favorably for plaintiffs on pretrial motions in race or gender discrimination cases than white judges (Boyd 2016). Another study also showed that white federal judges are less sympathetic to employment discrimination claims than minority judges—especially in pro se cases (Weinberg & Nielsen 2011). Yet another study concluded that plaintiffs in racial harassment cases have a 42.2% success rate with African American judges, a 20.6% success rate with white judges, and a 15.6% success rate with Hispanic judges (Chew & Kelley 2009). Moreover, in this study both white and African American judges were more likely to rule in favor of plaintiffs of their own race than for plaintiffs of other races.

Research examining the effect of judges' race on appellate panels in affirmative action cases revealed that African American judges support affirmative action plans 90% of the time compared with non-African American judges, who uphold such plans 80% of the time. The researchers also discovered that a judge's race has a larger panel effect than a judge's ideology (Kastellec 2013). Both results were consistent with an earlier study of affirmative action cases (Cameron & Cummings 2003).

Findings concerning the effect of judges' race on sentencing outcomes, however, are mixed. One study that examined sentencing in Pennsylvania from 1991 to 1994 found that African American judges are more likely to incarcerate offenders, but did not impose lengthier sentences than their white counterparts (Steffensmeier & Britt 2001). Another study comparing felony sentences imposed by African American and white judges from 1976 to 1978 in Michigan found that white judges were more likely to incarcerate than were African American judges (Spohn 1990). As in the previous study, however, this research found no effect of judges' race on sentence. A study of Latino judges in Texas similarly found that Latino judges were only slightly more likely to impose harsher sentences than their non-Latino counterparts (Holmes et al. 1993).

Research by Abrams et al. (2012) provides the strongest support for racial bias in judging, albeit using an indirect method. Using a large sample of trial court data from Cook County, Illinois, these authors concluded that African American defendants were sentenced more harshly than white defendants. The limitations of the data did not allow the authors to control for other factors, but they were able to demonstrate that, despite random case assignment, some judges exhibited a much bigger racial disparity than others. Their results suggest that some judges treat defendants of different races differently, but also that many judges do not.

As with other demographic variables, race seems to matter most when the issue of race lies at the center of the case. Although most of this research focuses on African American judges, concluding that only African American judges are motivated by race would be a mistake. For one thing, the data reveal that only certain kinds of cases come out differently when they are in front of African American judges. Litigants might be presenting their cases differently to an African American judge than to a white judge. Furthermore, the results might suggest that white judges react negatively to voting rights or affirmative action claims, at least as much as they suggest African American judges react positively to such claims.

Age. Age and experience are under-researched demographic factors. Fox & Van Sickle (2000) found that older judges were more apt to favor prosecutors in rulings in criminal cases. Another study found that younger judges were less sympathetic to plaintiffs in age discrimination cases than older judges (Manning et al. 2004). Indeed, the oldest judges were roughly twice as likely as the youngest judges to vote for plaintiffs in age discrimination cases. Although this study drew criticism on methodological grounds (Epstein et al. 2004), it adds to the emerging idea that judges' demographic characteristics induce empathy for claims litigants make that are related to those characteristics.

Employment history. Does pre-judicial service as a prosecutor predispose a judge to rule in favor of the defendant? Studies of the US Supreme Court suggest it does (Tate & Handberg 1991). Previous experience, however, might not be the kind of salient, personal characteristic that produces a clear attitude toward certain types of cases. Former prosecutors could be either more skeptical or more supportive of prosecuting attorneys who appear before them. The empirical evidence among the broader swath of state and federal judges is, in fact, mixed. Most studies conclude that judges who were former prosecutors were somewhat more likely to vote against the defendant (Nagel 1962). Robinson (2011), however, found no relationship between prosecutorial background and judges' decisions in criminal cases.

Summary

Overall, the pattern that emerges in these studies is clear, despite some wrinkles. The research does not support the view that there are widespread, consistent differences between judges based

on their background and beliefs. Judges, however, decide cases in which personal characteristics are salient in a manner consistent with their own personal characteristics. Judges, like all people, have political beliefs and so decide redistricting cases in ways that favor their beliefs. Devout judges tend to decide cases in ways that are consistent with their faith. Race, gender, and age can also make judges more sympathetic to discrimination claims by claimants who also share those characteristics with the judge. When it comes to the effect of demographics on judicial decision making, it is personal.

JUDICIAL ACCOUNTABILITY

Judges rarely get reversed on appeal, only occasionally get promoted, and hardly ever get removed from office. Each of these prospects, however, matters a great deal to judges, and several lines of inquiry suggest that these possibilities might affect how they decide cases (Baum 1997). Judges are also aware that their choices must be justified to the public and to their colleagues. Studies that show that the presence of a single female (or African American) judge influences her two male (or white) colleagues suggest a kind of accountability to judicial peers (Cameron & Cummings 2003, Farhang & Wawro 2004, Perisie 2005). As a general matter, the interaction between judges serving on appellate panels produces collegial pressure toward consensus (Fischman 2015), which also represents a form of peer accountability.

As a general matter, accountability is probably a healthy constraint on judges. The public needs to be able to accept the legitimacy of the judiciary, and holding judges accountable can further that acceptance. But judicial independence is also an essential feature of the third branch of government. Excessive attention to public opinion, the attitude of a reviewing court, or an upcoming election or reappointment can lead judges to make decisions that reflect personal concerns rather than the facts and law of the case before them.

Reversal Aversion

Review on appeal provides perhaps the primary mechanism for judicial accountability. The judiciary is hierarchical, and judges should, of course, follow binding precedent. Some scholars, however, have noted that in some instances, aversion to reversal can produce undesirable, strategic effects on judicial decisions.

To be sure, intermediate appellate review of trial court decisions is rare, and Supreme Court review of intermediate appellate court decisions is rarer still. In the federal courts, for example, approximately 11% of all cases filed in trial courts are appealed, and approximately 15% of appealed cases are reversed (Guthrie & George 2005). State courts exhibit a similar pattern (Liberato & Rutter 2003). Furthermore, although some judges might feel embarrassed by reversal or irritated by the extra work resulting from a remand, others may consider reversal a badge of honor if they disagree with prevailing precedent. Given that reversal occurs so rarely, it is small wonder that studies have difficulty identifying any effect that the prospect of reversal might have (Klein & Hume 2003, Songer et al. 2003).

A careful analysis by Epstein et al. (2013) nevertheless found an indirect measure of reversal aversion among federal district court judges. The authors discovered that although Republican and Democratic judges did not decide cases differently, the politics of the appellate judges who would review their cases influenced their judgment. This led the authors to conclude that district court judges follow precedent more so than their own attitudes, whereas appellate judges are free to express their political preferences with only a minimal fear of reversal by the US Supreme Court (which takes few cases). Although alternative accounts might explain the results (district

court judges might focus more on facts, which may be less amenable to political influence than judgments of law), other studies support the conclusion that avoiding reversal matters to judges.

Schanzenbach & Tiller (2006) demonstrated strategic behavior among federal district court judges to avoid reversal in sentencing decisions. Under the Federal Sentencing Guidelines before *Booker v. United States* (2005), federal judges could impose sentences below the range prescribed by the Guidelines only if they either identified mitigating factors or concluded that as a matter of law the Guidelines did not address the defendant's situation. The two findings produce the same result for a defendant (a shorter sentence), but the former is reviewed on appeal as a determination of fact reversed only for clear error; the latter gets more scrutiny on appeal under an abuse of discretion standard. The authors found that Democratic district court judges imposed shorter sentences overall than their Republican counterparts. And in circuits in which Democratic judges faced appellate panels that were mostly Republican, they strategically found facts to reduce sentences, so as to avoid more searching review on appeal. Randazzo (2008) also found that trial judges seemed to conform to the political ideology of the appellate judges who review their decisions, but Boyd & Spriggs (2009) failed to find evidence that judges strategically defend their decisions from ideologically different appellate panels.

The overall picture painted by these studies suggests a healthy effort among trial judges to conform to the law of the jurisdiction. But the research also reveals that they engage in a measure of strategic behavior to shield their decisions from appeal.

Auditioning

Conforming to the ideology of a circuit is defensible, but strategically deciding cases to obtain a promotion is troubling. In one Annenberg survey, 75% of the public stated that they believe “a desire to be promoted to the next higher court would affect a judge’s ability to be fair and impartial when deciding a case” (Jamieson & Hennessy 2007, p. 900). Most judges have low odds of promotion to a higher court and remain in the judicial position to which they were initially appointed or elected throughout their careers. Nevertheless, the decisions of some judges appear to be influenced by the possibility of promotion. In particular, auditioners appear to impose longer sentences and otherwise avoid the appearance of being soft on crime when contrasted with nonauditioners (Epstein et al. 2013). Similarly, federal district judges with higher probabilities of being promoted to the court of appeals were less likely to rule that the new federal sentencing guidelines were unconstitutional than those whose prospects for promotion were poor (Cohen 1991). State-court judges exhibit similar tendencies (Brace et al. 1999).

Electoral Accountability

Though often criticized, the election of judges is arguably a defensible mechanism of accountability. The proximity of an election, however, is an inappropriate influence on judicial decision making. Even though reappointment and reelection are overwhelmingly likely (Aspin 2007), some studies show that judges’ decisions nevertheless are influenced by impending retention elections or reappointment. For example, a study of Pennsylvania judges facing retention elections at 10-year intervals found that they imposed more severe sentences shortly before (and even after) their retention elections than they did at other times during their tenure (Huber & Gordon 2004). One other study indicates judges facing retention elections are less favorable to capital defendants’ efforts to overturn their sentences (Blume & Eisenberg 1999). The effect is not confined to elections. Judges approaching consideration for reappointment exhibit similar behavior (Shepherd 2009).

Worse still is the concern that judges are affected not by their elections but by their donors. In extreme circumstances, judges must recuse themselves in cases involving parties who have made large donations to their election campaigns [e.g., *Caperton v. Massey Coal Co.* (2009)]. Regrettably, there is evidence of a link between contributions and judges' decisions (Cann 2007). Subtle effects of campaign contributions are apt to go unnoticed, however. Kang & Shepherd (2015) show exactly this kind of untoward effect in state supreme court elections. Using multiple measures, these authors show that donations from a political party correlate with judicial decision making, especially among Republican judges. Their results suggest that it may be difficult for judges to ignore partisan efforts that enable them to retain their positions (Shepherd & Kang 2011). Similarly, Cann (2007) found that in Georgia, an attorney's campaign contributions influence judicial decision making in the attorney's cases before that judge.

Summary

Overall, facing one or more forms of accountability is a healthy and necessary part of the judicial role. But in some settings, judges strategically avoid accountability by how they decide cases. The prospect of promotion or the pressure of elections can also distort the decisions of some judges in ways that fall short of a judicial ideal.

INTUITIVE REASONING IN JUDGES

Do judges rely on simple mental shortcuts to make decisions? Psychologists know that most people manage complex environments by simplifying the decision-making landscape. Judges do not typically describe their jobs that way, however. In particular, people often rely excessively on intuitive strategies to make decisions, even in settings in which such strategies lead to mistakes (Kahneman 2011). Empirical research suggests that judges are no different.

Judges Are Intuitive Thinkers

Judicial performance on the cognitive reflection test (CRT) provides a prime demonstration that judges often rely too heavily on intuitive reasoning (Guthrie et al. 2007). The CRT is a simple three-item test, reprinted below, which pits intuitive reasoning against deliberative reasoning (Frederick 2005, at p. 27):

A bat and a ball together cost \$1.10. The bat costs \$1.00 more than the ball. How much does the ball cost?

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

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?

Each item in the CRT suggests an intuitive answer (10 cents, 100 minutes, 24 days) that is wrong. Answering correctly (5 cents, 5 minutes, 47 days) requires suppressing the intuitive answer and thinking more carefully. The questions themselves are not difficult, but most well-educated adults get most of them wrong.

Although simple, the CRT is a surprisingly powerful test. People who tend to think deliberatively on the CRT tend to think deliberatively in other domains (Frederick 2005). People who

score well on the CRT are wealthier and even live longer than those who do not (Weber & Johnson 2009). They are deliberative thinkers who make better decisions.

Are judges, as a group, particularly deliberative? No. Judges, like most adults, get most of the CRT questions wrong (Guthrie et al. 2007). Although judicial tasks commonly require suppressing intuitive reactions and deliberating carefully, most judges do not naturally stop to suppress their intuition. Most of them rely on their first instinct on the CRT, even though it is wrong. Unlike engineers (who score well on the CRT; Frederick 2005), judges seem to be intuitive thinkers.

Judicial Intuition on Legal Problems

One might forgive judges for not performing well on math problems, but judges also rely too heavily on intuitive responses to questions in legal domains. In one study, 160 federal judges evaluating a hypothetical case neglected statistical evidence in favor of intuition in the assessment of negligence (Guthrie et al. 2001). The stimulus materials for the study converted a widely studied problem of statistical neglect into a legal setting using the foundational case for the doctrine of *res ipsa loquitur*, *Byrne v. Boadle* (1863). In typical studies of this type, decision makers fail to recognize that a rare event is more likely to be the unlikely result of a common occurrence than the likely result of an uncommon occurrence (Kahneman & Tversky 1973). Judges in the study assessed the likelihood that a warehouse was negligent for an accident involving a barrel that injured a bystander. The materials reported that an investigation had revealed that when the warehouse is careful, accidents occur one time in 1,000, but that when the warehouse is negligent, accidents occur 90% of the time. The materials also indicate that the defendant is negligent only 1% of the time. These statistics indicate that the likelihood that the accident was the result of negligence was 8.4%. Over 40% of the judges, however, concluded that negligence was more than 75% likely, with a similar percentage (correctly) concluding it was less than 25% likely. Even in a legal setting, about half of the judges relied on their intuitive reaction that the defendant was negligent, even though proper mathematical analysis indicates otherwise.

The results of the barrel problem reflect the development of doctrine in case law. According to the traditional formulation of the doctrine of *res ipsa loquitur*, a court can conclude that if the likely product of a defendant's negligence is an accident, then the fact of that accident supports the conclusion that the defendant is negligent (Am. Law Inst. 2009). This is a logical fallacy that ignores the overall rate of negligence. Nevertheless, the fallacy persisted in the case law for over a century, as one court after another failed to recognize it (Kaye 1979). Ultimately, the American Law Institute (2009, chapter 17) identified and corrected the logical fallacy in the *Restatement (Third) of Torts*. In the study, many judges ignored the base rate statistics, just like the judges who developed the doctrine.

Related work shows that judges suffer from another logical fallacy that arises from an excessive reliance on an intuitive response: confirmation bias. Confirmation bias is an effort to seek out information that is consistent with one's prior beliefs, while ignoring or avoiding information that could refute them (Wason 1968). In the classic demonstration of the phenomenon, Wason showed people four cards displaying either a 3, 8, e, or p and asked them to assess the statement that "if a card has an odd number on one side, it has a vowel on the other side" by turning over as few cards as possible. Most people correctly choose the 3, but then incorrectly choose the e, even though the e cannot provide any evidence as to the correct hypothesis. The p card is correct, because an odd number on the other side of the p would falsify the statement. Judges are no different, as 90% solve this problem incorrectly (Wistrich & Rachlinski 2013). To put these results in perspective, it should be noted that doctors (Wallsten 1981), lawyers (Wistrich & Rachlinski 2013), and arbitrators (Helm et al. 2016) commit the same mistakes. Furthermore, even when judges review a version of

the problem set in the context of a discovery request in a civil lawsuit, they perform only slightly better (Wistrich & Rachlinski 2013). Judges ordered compliance with discovery requests that could not be relevant to the underlying case while denying requests that could have undermined one party's theory of the case. Untested, but perhaps more worrisome, is whether judges deliberately avoid information that might undermine their first impressions of a case.

Judges commit other logical errors as well. In one study in which judges evaluated an alleged case of discrimination against a Muslim employee named Dina, most judges ranked the likelihood that the employer both "actively recruited a diverse workforce" and also "unlawfully discriminated against Dina based on her Islamic religious beliefs" as higher than the likelihood of either of the two events in isolation (Guthrie et al. 2009). In another experiment, judges reacted differently to the probability that a piece of forensic evidence matched that of the perpetrator at 0.1% as being different than 1 in 1,000 (Rachlinski et al. 2013b). Judges also pay special attention to the frame of a decision—that is, whether it involves gains or losses (Guthrie et al. 2001, 2009). For example, a study of bankruptcy judges showed that their willingness to approve a reorganization plan depended upon whether the materials described the same plan in terms of gains as opposed to losses (Rachlinski et al. 2006).

Overall, the research shows that, at least in the evaluation of hypothetical legal cases, judges rely heavily on intuitive mental shortcuts in a wide range of contexts. Viscusi (1999) also reached similar conclusions in his study of judges. He found that judges relied on many simple, misleading heuristics in evaluating legal fact patterns. Whether judges use the same heuristics in actual cases is less certain, but experimental studies show heavy reliance on heuristics by judges.

Judicial Intuition: Studies of Actual Cases

Even apart from these experimental demonstrations using hypothetical cases, judges often embrace simple decision rules when facing complex choices. Decades ago, psychologists Ebbesen & Konečni (1975) demonstrated that judges rely excessively on the most salient information when setting bail and sentencing criminal defendants. In both situations, judges face a complex array of factors that must be reduced into a single numeric decision. The judges that Ebbesen & Konečni (1975) studied reported that when setting bail they considered the crime charged, the defendant's past criminal history, the defendant's ties to the community, the defendant's employment history, whether the defendant seemed likely to commit more crime while out on bail, and the probability that the defendant would later appear for trial. All are sensible factors. But in evaluating both hypothetical cases and real ones, judges relied almost exclusively on the prosecutor's recommendation. An analysis of criminal sentences produced a similar result (Ebbesen & Konečni 1982). More recent research on judges in Britain (Dhami 2003) and Spain (Fariña et al. 2003) uncovered a similar pattern; that is, although judges claim that they consider a wide array of factors, they appear to rely almost exclusively on prosecutorial recommendations.

To be sure, these studies come with an important caveat: Prosecutors and judges likely believe the same factors to be critical in setting bail. In both studies, these factors correlated with both judges' and prosecutors' decisions. That said, the correlation between the prosecutors' recommendations and judges' sentences is so strong that it is hard to imagine that the recommendation has no independent influence. Ebbesen & Konečni (1975) used carefully controlled hypothetical decisions to isolate the effect of prosecutors' recommendations on judges. What is more, judges claimed that they were ignoring the prosecutors, which is difficult to believe, given how closely their decisions coincide with the recommendations. Judges, like many people, seem to be managing a complex environment by relying on simple mental shortcuts. And they seem unaware of what they are doing.

An analogous study of judges in a completely different setting provides similar support for the conclusion that judges simplify their decision making. Beebe (2006) found a similar pattern in his assessment of the decisions made by judges in trademark cases. Beebe found that although the federal courts follow a variety of multifactor tests to assess consumer confusion in trademark cases (incorporating over eight factors), judges rely on only two (similarity and proximity of the commodities in question). As with the research on bail and sentencing decisions, judges realize that they are supposed to consider other factors as well, and state that they have done so. Beebe's systematic empirical assessment shows that most of the factors do not predict the outcome. Beebe found similar results with a study of fair use in copyright law (Beebe 2008). Awards of attorneys' fees likewise depend upon the weighing of numerous factors, many of which do not have any effect on judges' actual awards (Eisenberg & Miller 2004). It seems that judges face factor overload, which induces them to rely on simpler, more intuitive decision-making rules (Levy 2013).

Stressful conditions could induce judges who would otherwise adopt complex approaches to rely on simple strategies. A widely discussed study of Israeli parole judges found that these judges were far more likely to make the tough decision to grant parole immediately after they ate breakfast, a mid-morning snack, or lunch and were likely to make the easier, default decision to deny parole near the end of each session (Danziger et al. 2011). The authors also provide an alternative account of the results. They suggest that judges might have implicit quotas as to how many prisoners should be granted parole in any particular session. As the quota fills, they begin to deny most of the remaining applications. In either case, the time of day of the case mattered enormously, with the chances of parole dropping from about 69% at the beginning of the day to less than 20% before lunch.

A related study on sequential decision making by Chen et al. (2016) suggests that a similar effect occurs in immigration judges. This study showed that when judges decide a few cases in a row the same way, they begin to feel that the next case should come out differently. This phenomenon, known as the gambler's fallacy, relies on a misperception of chance. Even though runs of three or more of the same outcome are common in a random binary sequence in which the odds of either outcome are 50%, such runs seem unnatural and hence unlikely. Judges seem to believe that deciding a sequence of unrelated cases in a way that favors the same side is also unnatural, and actively avoid deciding cases in a way that favors the same side several times in sequence.

The Israeli study, though presenting dramatic results, also tells a cautionary tale: Failing to fully grasp the environment in which judges' decisions are made can cause misinterpretation of results. This is a potential vulnerability of archival studies. In this instance, the researchers failed to account fully for the fact that the parole judges triaged their calendars by calling the cases in which the prisoner was represented by counsel earliest in each session; this provided an intuitively plausible alternative explanation for the observed results because prisoners who are represented by counsel fare twice as well as those who are not (Weinshall-Margel & Shapard 2011).

Overall, the evidence strongly indicates that judges will, in some cases, rely on simple intuitive strategies. The data on judges and the CRT show them to be intuitive thinkers on the whole. They rely on intuitive answers to solve legal problems as well, even when intuition leads them astray. And in real cases, the research suggests a pattern of reliance on rules of thumb that simplify their decisions—even when they claim (and believe) that they are not doing so.

Anchoring in Judges

One mental shortcut that has received sustained attention is anchoring. When making numeric judgments, people commonly rely on numeric reference points as anchors (Tversky & Kahneman

1974). This heuristic is often extremely useful, as initial numeric reference points are often reliable, but people rely heavily on numeric anchors, even when it is not sensible to do so (Ariely 2009). Research indicates judges do so as well.

A variety of factors have been demonstrated to anchor judges' numeric assessments (Rachlinski et al. 2015). Judges evaluating hypothetical cases have relied on such anchors as a damage cap, a damages demand, a damage award in an unrelated case, the sentence imposed in the preceding case, a sentence urged by the prosecutor, and a sentence recommended by a probation officer (Rachlinski et al. 2015). Similarly, bankruptcy judges' choice of interest rate was affected by exposure to the original contract interest rate, something doctrine explicitly requires them to ignore (Rachlinski et al. 2006).

Anchoring can generate ridiculous outcomes. In another experiment, judges awarded eight times as much to a hypothetical plaintiff who mentioned that she had watched a court TV show in which a similar plaintiff received an award of \$415,300 (Guthrie et al. 2009). In another study, municipal court judges fined a nightclub three times as much when its name (after its street address) was Club 11,866 rather than Club 58 (J.J. Rachlinski & A.J. Wistrich, unpublished manuscript). Finally, in an experiment involving German judges, even a roll of the dice affected judges' sentencing decisions in a hypothetical case (Englich et al. 2006).

Disentangling erratic from reasonable effects of anchors in real cases is challenging. As noted above, judges are heavily affected by prosecutors' recommendations (Ebbesen & Konečni 1975). Prosecutors' recommendations are not randomly selected, of course; they reflect the same meaningful factors that judges say they consider (Leifer & Sample 2010). Sentencing recommendations that arise from erroneous information, however, influence sentences just as much as accurate recommendations (Bushway et al. 2012). A study of Taiwanese judges likewise shows that misleading recommendations can influence judges in civil cases (Y. Chang, K. Chen, C. Lin & Y. Liu, unpublished manuscript). These studies indicate that recommendations anchor judicial decisions, even when they are misleading. The research on real cases thus complements the conclusions of the experimental studies.

Usually an anchor consists of a single number—such as a damage cap—but other forms of numeric context can also wreak havoc with judges' qualitative assessments. One study showed that judges imposed a sentence approximately 40% shorter on the same defendant for the identical crime when asked to sentence in months rather than years (Rachlinski et al. 2015). Although that experiment was based on reactions to hypothetical cases, it mirrors the results of a change in law in Finland (Lappi-Seppälä 2001). When judges switched sentencing from months to days, sentences dropped by 25% (although other reforms were also introduced at the same time). A study of Canadian judges showed that judges imposed sentences (for second-degree murder) that tended to be multiples of five or (to a lesser extent) even numbers, even though every integer is available to judges (Jones & Rankin 2015). Presumably the degree of desert is a continuous parameter, and yet judges fixated on multiples of five.

Oddly enough, judges not only rely on numeric anchors but also exhibit the opposite effect as well—contrast effects. Contrast effects are perhaps best explained by example. In one experiment, judges' assessments of the relative credibility of two well-qualified experts, a psychiatrist and a psychologist, were altered by the addition of a poorly qualified psychiatrist (Rachlinski et al. 2013b). Those judges who read about the poorly qualified psychiatrist found the well-qualified psychiatrist more credible than those who had not also evaluated the poorly qualified psychiatrist. The addition of the unqualified psychiatrist made the qualified psychiatrist look better in contrast. Archival analysis suggests that the same phenomenon manifests itself in juries' damage awards, but this same study did not uncover this effect for judges' damage awards in bench trials (Eisenberg et al. 2002).

One recent study, based on data reported by the Pennsylvania Sentencing Commission, tested whether contrast effects or anchoring influences criminal sentences in actual cases (Leibovitch 2016). This study identified instances in which trial judges had recently sentenced unrelated defendants—more or less duplicating an experimental study that demonstrated that preceding cases can anchor judges' sentences (Rachlinski et al. 2015). Judges were influenced by their sentence in an immediately preceding case but expressed a contrast, rather than anchoring, effect. Judges who sentenced defendants for very serious crimes during the preceding nine months imposed sentences that were 25% shorter for less serious crimes than those imposed by judges who had not recently sentenced a serious offender. Scholars have long noted that anchoring and contrast effects are conflicting phenomena (Sherif et al. 1958). In some instances, comparisons produce anchoring, and in others, contrast effects. The study of sentencing shows that the pattern of sentencing in Pennsylvania, at least, produces contrast, rather than anchoring, effects. The experimental research and a handful of other studies, however, suggest that in other settings, anchoring will occur. Of course, neither is an appropriate influence, and assessing when anchoring occurs and when a contrast effect occurs will require more research.

These studies show reliance on simple heuristics, but they do not paint a clear portrait. Judges in experiments show both contrast and anchoring effects; judges in one study show contrast effects in criminal sentences, whereas judges in another study do not show the same effect in civil damage awards. Nevertheless, judges seem to produce erratic results that are consistent with contrast and anchoring effects. The overall pattern indicates that judges have difficulty reliably converting qualitative information into a single quantitative number and rely heavily on anchoring.

Summary

Research on judges strongly points to the conclusion that judges rely heavily on intuitive reasoning to evaluate legal disputes. Both experimental and archival studies show that judges use simple mental shortcuts to guide how they think about legal materials. The same research also suggests that judges do not improve with experience or specialization (Guthrie et al. 2009, Rachlinski et al. 2006). Indeed, experience might induce judges to adopt mental shortcuts that they did not use when they were new judges. One study of judges who decide many patent cases, for example, finds that they begin to develop stylized patterns of decision making that disfavor the patentee (Lemley et al. 2014). Doubtless judges can and do engage in more deliberative, complicated analysis at times, but the evidence suggests that their reliance on intuitive reasoning leaves them vulnerable to errors in judgment.

FACTORS OUTSIDE THE RECORD: INADMISSIBLE EVIDENCE, EMOTION, AND INVIDIOUS BIASES

The kinds of heuristics that judges seem to use create the potential for other serious problems. Do judges decide entirely on the record? Do they rely on emotional reactions and sympathy instead of logic and the law? And worse yet, do they favor in-groups, as the data on judicial demographics suggest? Unfortunately, research provides evidence for all of these effects.

Deciding on the Record Alone: Inadmissible Evidence

If judges rely on simple mental shortcuts when deciding cases, then they likely will face difficulty limiting the scope of their decision making to the record. Judges must found their decisions on the record and only on the record. They must set inadmissible evidence aside. Jurors find this

task difficult; dozens of mock jury studies show that no reliable mechanism exists to expunge inadmissible evidence from jurors' minds (Stebly et al. 2006). Some studies find that instructions to disregard inadmissible evidence are effective, others find that they have no effect, and still others show that such instructions induce jurors to pay more attention to the evidence they are instructed to disregard (Wistrich et al. 2005). Studies comparing how judges assess damages to how jurors do so indicate that judges use similar strategies and have similar weaknesses (Wissler et al. 1999). Whereas information can be kept from a jury, however, judges cannot shield themselves from inadmissible evidence. They are both the gatekeepers of evidence and the decision makers.

Testing the effect of inadmissible information has required an experimental approach. Researchers have conducted studies in which judges review materials that include or omit inadmissible information (Landsman & Rakos 1994, Wistrich et al. 2005). If judges can ignore inadmissible evidence, then those judges exposed to such evidence should make roughly the same decisions as those judges not exposed to the inadmissible evidence.

In one such study, a group of trial judges in Arizona analyzed a hypothetical bench trial in a criminal case in which a college coed accused a fellow student of sexual assault during a fraternity party (Wistrich et al. 2005). The defendant admitted to having intercourse with the complainant but asserted that the intercourse was consensual. The materials indicated that the complainant contacted police immediately after the incident and had bruising consistent with a sexual assault. Half of the judges read this version and then reached a verdict. Among these judges, 49% convicted. The other half also read that the defendant attempted to introduce testimony that the complainant's roommate stated that the complainant "liked to loosen her inhibitions with a few beers too many and have rough sex with the first guy she saw." Such testimony is inadmissible under Arizona's rape-shield statute. Most of the judges ruled the testimony to be inadmissible, but the conviction rate among these judges dropped to 20%. These judges did not truly suppress the inadmissible testimony; rather, they relied on it.

Other studies found a similar inability to disregard inadmissible evidence in other contexts. For example, judges could not ignore a discussion protected by attorney-client privilege in a civil case, the past criminal conviction of a civil defendant, discussions that occurred during a settlement conference, and statements made by a criminal defendant that a prosecutor had agreed not to use in a plea agreement (Wistrich et al. 2005). Yet another study found that judges considered subsequent remedial measures in a products liability case, even when informed they were inadmissible (Landsman & Rakos 1994).

Although judges do not ignore inadmissible confessions, the influence of such a confession varies depending on the circumstances (Rachlinski et al. 2013b). One experiment demonstrated this variability using a hypothetical bench trial in a prosecution for an armed bank robbery. The judges' willingness to disregard evidence obtained in violation of the defendant's constitutional rights depended on the nature of the crime and the nature of the constitutional violation. Judges relied on the inadmissible confession to convict the defendant when the crime was more serious and tended to ignore the confession when the police misconduct was more egregious.

In one setting—that of probable cause determinations—research suggests that judges are able to ignore information that they are not supposed to consider (Rachlinski et al. 2011). One might expect that judges assessing probable cause in search-and-seizure cases suffer from hindsight bias. That is, they might base their assessments of whether probable cause exists largely on whether the search produced incriminating evidence. A series of experimental studies using hypothetical cases, however, found that judges make the same decisions regarding probable cause in foresight (in assessing a warrant application) as they do in hindsight (assessing whether evidence obtained pursuant to an exception to the warrant requirement lacked probable cause). This research suggests that judges are not relying on probability judgments when assessing probable cause, but on the case

law. In other contexts, however, judges have trouble ignoring post-event knowledge (Anderson et al. 1993; Guthrie et al. 2001, 2009).

Judges also seem unable to disregard inadmissible information from other extraneous sources. Well-founded concerns that jurors might be improperly influenced by media accounts of cases (Daftary-Kapur et al. 2014, Moran & Cutler 1991, Otto et al. 1994) have prompted jury instructions (9th Cir. Model Jury Instr. 1.7, 2007) and even sequestration (Strauss 1996). Recent evidence suggests that judges share that susceptibility (Lim 2015, Lim et al. 2015).

The human brain is not designed to ignore information. Although legal scholars debate whether judges have a special ability to disregard irrelevant information, the conclusion that “[n]ature does not furnish a jurist’s brain with thought-tight compartments to suit the convenience of legal theory, and convincing evidence does leave its mark” (Maguire & Epstein 1927, pp. 1116–17) seems like the correct one.

Emotional Decision Making in Judges

Just as judges seem to believe they can ignore inadmissible information, judges commonly deny that emotion influences their decisions. In her confirmation hearing, for example, Justice Elena Kagan was asked if she agreed that “law is only 25 miles of the marathon and emotion is the last mile.” She rejected the assertion outright by claiming “it is law all the way down” (Senate Comm. Judic. 2010). In contrast, Justice Robert Jackson likened the image of dispassionate judges to that of “Uncle Sam, Santa Clause, the Easter Bunny, and other fictional characters” [*United States v. Ballard* (1944)]. Justice Jackson’s comments are a rare admission, as most judges understand that society expects judges to be dispassionate (Maroney 2011). Nevertheless, judges clearly have emotional reactions to litigants. One study of trial court judges in California showed that reviewers who knew nothing about the case and could not hear the proceedings could nevertheless accurately assess a judge’s reaction to a witness through the judge’s nonverbal expressions (Blanck et al. 1985).

A series of experiments with hundreds of judges from numerous jurisdictions concluded that emotions influence how judges interpret law when evaluating hypothetical cases (Wistrich et al. 2015). In one study, judges interpreted a medical marijuana statute more favorably for a defendant who was dying from bone cancer than for a 19-year-old defendant who was suffering from seizures. In another, judges interpreted the law on forgery more favorably for an undocumented immigrant who had entered the United States to earn money for a sick daughter than for an undocumented immigrant who had entered the United States to hunt down a rogue member of a drug cartel. Judges were more likely to rule a city jail’s blanket strip-search policy was per se unconstitutional when the lead plaintiff was a college coed protestor compared with a male armed robber. Bankruptcy judges indicated they would discharge more debt of a debtor who borrowed to help an ailing parent than one who borrowed to go on vacation, even though the relevant law does not distinguish based on the source of the debt. Notwithstanding Justice Kagan’s assertions, emotion seems to be some portion of the marathon.

Another experimental study of federal judges showed similar effects. Spamann & Klöhn (2016) asked federal judges to review a complex set of materials that presented an issue of statutory interpretation. They varied two factors: The litigant was either highly sympathetic (the victim of a war crime) or highly unsympathetic (a war criminal), and the precedent either favored or opposed the litigant’s position. They found that the litigant’s identity heavily influenced the judges, but the precedent had no effect. The judges thus ignored the relevant legal parameters that should have influenced their judgment while being moved by the emotional parameters that they should have ignored. Judges’ feelings about the nature of the crime also influence their judgment.

Other studies show a similar influence of emotion on judges' judgment. One study showed that judges set higher bonds for defendants charged with sex offenses than for defendants charged with non-sex offenses of equivalent statutory offense level (Beattey et al. 2014). Such crimes may not be worse according to an abstract hierarchy of gravity, but they feel that way to judges, which provokes a more severe sentence. Another study showed that these feelings can be manipulated as well (Rosenblatt et al. 1989). These authors asked municipal court judges to assess a hypothetical bail hearing for a prostitute after writing a short essay about either what they had for lunch or what happens to their bodies after they die. The latter task induces a fear of death in most adults that leads people to make decisions that more firmly uphold moral values. Judges did in fact react to the task, assigning much higher bail to the prostitute after contemplating their mortality.

A recent study revealed a particularly invidious influence of emotion on judges. Eren & Mocan showed that judges in Louisiana sentence juvenile offenders to harsher punishment in the days following an unexpected loss by the Louisiana State University football team (O. Eren & N. Mocan, unpublished manuscript). The authors carefully studied years of court records in juvenile cases, matching them to outcomes of the highly popular team. Compared with weeks in which the team did not play, won, or lost when a loss was expected, the authors found that judges imposed longer sentences on African American juveniles after an unexpected loss. The result should be taken with some degree of caution. Many of the sentences were suspended—which the authors did not factor into their analysis. Their conclusion that the effect is equally strong throughout the whole week after a game is also puzzling, as is the fact that it influences sentencing of only African Americans. But the result is disturbing and bears close attention.

Related lines of inquiry show that physical attractiveness influences judges. According to one study based on in-court observations, attractive defendants received shorter sentences and were less likely to be incarcerated than unattractive defendants, although there was no difference in the rate at which they were found guilty (Stewart 1980, 1985). Other researchers studied attractiveness in the context of release on bond or detention before trial (Downs & Lyons 1991). They found, again based on in-court observation, that attractive defendants were required to post lower bonds than unattractive defendants when charged with a misdemeanor, but both sets of defendants were treated equally when charged with a felony. A study of British magistrates likewise found that more attractive defendants are less likely to be convicted (Block 1991). Outside the criminal context, one study found that more attractive litigants in small-claims court tended to receive better outcomes (Zebrowitz & McDonald 1991). Experiments on judicial susceptibility to beauty bias are few but have not revealed an effect (Guthrie et al. 2009).

Salient events can change how judges react to cases as well. Sisk & Heise (2012) found that the association between Muslims and terrorism affected how judges decided religious freedom cases. These authors found that up until the mid-1990s, religious accommodation and free exercise claims among Muslims and non-Muslims were equally likely to succeed. After the growing public perception of an association between Muslims and terrorism that began in the mid-1990s, however, the success rate of such claims by Muslims plummeted, even as those by non-Muslims remained constant. Sounding a more optimistic note, a study by McCall (2003) concluded that the Anita Hill–Clarence Thomas hearings changed how state supreme court justices viewed sexual harassment cases. Before the hearings, male judges were less accepting of such claims than female judges, but the salience of the issues raised by these hearings seemed to close that gap.

To be sure, judicial reliance on emotion in decision making can be defensible (Maroney 2011). Judges should temper their application of law and logic with expressions of compassion and empathy. Indeed, one set of studies finds that judges seem to largely ignore apologies in both civil and criminal cases, making the judges seem overly dispassionate (Rachlinski et al. 2013a). The studies described in this section, however, go well beyond a sensible level of compassion. No one

can defend taking a football loss out on juveniles, setting lower bail for more attractive litigants, or treating Muslim litigants differently after 9/11. Nevertheless, these studies show judges to be vulnerable to several such untoward influences.

Invidious Biases Based on Group Membership

Sisk & Heise's (2012) study of how Muslims fare in the courts suggests a troubling willingness among judges to react negatively to litigants perceived as outsiders. This study provides evidence of what is perhaps the most potentially pernicious way in which judges fall short of their ideal. The goal of equality before the law depends upon judges setting aside their biases. Judicial codes of ethics and judicial oaths demand that judges disregard prejudices based on race, gender, ethnicity, or other outsider status. Prejudice and implicit bias are widespread among most adults, and although judges hold themselves to a very high standard in this regard, research provides evidence of such influences on judges as well.

Foreign litigants. One study in which judges from Minnesota evaluated a case of environmental pollution shows just how easy it is for judges to react negatively to outsiders (Rachlinski et al. 2015). The materials in this study described a defendant in a civil case who had deliberately dumped highly toxic chemicals into a lake on the plaintiff's property, severely injuring him. The materials asked the judges whether the case merited punitive damages (almost all judges thought it did) and if so, how much would be appropriate. The materials indicated that the plaintiff was a Minnesota resident. For half of the judges, the defendant was also from Minnesota, and for the other half, the defendant was from Wisconsin. The results were dramatic. Although the defendant's home state was of no legal significance, the median award rose from \$1 million against the Minnesota defendant to \$1.75 million against the Wisconsinite. The judges severely punished the Wisconsinite for being an outsider. The study found similar, albeit smaller, effects in New Jersey (with Pennsylvania as the foreign jurisdiction) and Ohio (with Michigan as the foreign jurisdiction). The results are also similar to those of an archival study of judicial decisions in tort cases (Helland & Tabarrock 2002).

Litigant gender. Gender matters to judges. Two areas of law, child custody and sentencing, show broad and pervasive disparities favoring women (or disfavoring men). Researchers have long recognized a deeply held trend in custody decisions—women usually obtain custody of children upon divorce (Kunin et al. 1992). "Throughout most of the . . . [Twentieth] Century the mother was presumed to be the preferred custodian of the children" (Maccoby & Mnookin 1992, p. 7). Contemporary reviews of the research suggest that this has changed little over time (Paradise 2012). A recent experimental study confirms favoritism among judges for female litigants in child custody disputes (J.J. Rachlinski, A.J. Wistrich & C. Thöni, unpublished manuscript). Disparity in outcome, however, cannot alone support a conclusion that judges favor female litigants in custody disputes. A big disparity that arises from deeply held cultural mores can hardly be laid entirely at the feet of judges. Divorcing parties might prefer that outcome, or women might litigate much more vigorously over custody, perhaps compromising on other aspects of the divorce to ensure that they retain it. The research in this area does not truly single out the role judges play in producing the large disparity in outcomes. One archival study also shows that although women fare well in initial custody hearings, the preference for mothers is smaller in later proceedings (Kunin et al. 1992). This result suggests that as judges attend more closely to cases, they rely less on cultural stereotypes.

Judges also disfavor men in sentencing. According to Etienne (2010, at p. 73), "the sentencing disparities based on gender are among the most visible and persistent sentencing disparities in this

country.” As with custody, a widespread disparity at the end point is not necessarily attributable to sexism among judges. Male and female offenders might have different background characteristics, on average. Studies of drug cases that control for offender characteristics, however, show sentences are much more lenient for women (Albonetti 1997, Spohn 2013). Even so, prosecutors and other court officials might also charge male and female defendants differently or make different recommendations for male and female offenders. Another study shows that women are treated more favorably throughout the charging and sentencing process, resulting in male offenders receiving sentences that are 63% longer than those for female offenders (Starr 2012).

Studies of downward departures under the Federal Sentencing Guidelines provide perhaps the best methodology for controlling factors other than race in sentencing. The Guidelines require judges to sentence within the Guidelines or explain their departures. The Guidelines themselves incorporate a wide range of factors that neatly control for case and offender characteristics. Departures represent a highly discretionary use of judicial power such that patterns in departures might reflect systematic favoritism (or disfavoritism) for certain kinds of litigants. Several studies show that women draw more downward departures than men (Mustard 2001, Schanzenbach 2005). The result suggests that judges truly treat women differently in sentencing, even when virtually all other factors are controlled.

Sentencing in juvenile cases reveals a more complex pattern. One study of juvenile sentencing in Hawaii replicates adult sentencing in that female juveniles received more lenient treatment (MacDonald & Chesney-Lind 2001). The result, however, was limited to early stages of the proceedings. Girls convicted of more serious crimes were treated, if anything, more harshly than their male counterparts. A study of Texas juvenile offenders replicated and extended the results, finding that girls who had been victimized, ironically, received even longer sentences (Espinosa & Sorensen 2016). This reversal might reflect the nature of juvenile court as more focused on rehabilitation. Judges might treat girls in a more paternalistic fashion than boys, feeling they would benefit from more time in a highly supervised environment. In any event, the pattern is clear: Judges do not treat males who come before them in the same way as females.

Women do not always fare better than men in court, of course. One experiment showed that judges award less for a female decedent than an otherwise identical male decedent (J.J. Rachlinski & A.J. Wistrich, unpublished manuscript). This result might reflect real differences in lifetime earnings between men and women, but it still represents an unsettling disparity in judicial treatment of male and female litigants.

Litigant race. Disparities in criminal justice outcomes are among the most significant racial disparities in American society. A meta-analysis of 71 studies of sentencing shows that African American defendants consistently receive harsher sentences than white defendants (Mitchell 2005). Studies of bail also show wide differences that are hard to explain without reference to race (Ayres & Waldfogel 1994). As with differences between the sexes, however, even a widespread disparity does not necessarily indicate that race influences judges’ decisions. Research indicates that prosecutors and pretrial services or probation officers treat African American defendants more harshly than white defendants, which will invariably affect bail and sentencing decisions. But some evidence suggests judges themselves are at least partly responsible for sentencing disparities (Ruback & Vardaman 1997, Spohn et al. 1987). One study, for example, found that judges were more receptive to evidence of remorse offered by white defendants than African American defendants (Everett & Nienstedt 1999). As with the role of gender in sentencing, studies of sentences under sentencing guideline systems provide a means of focusing on the judge. Federal judges issue far more downward departures for white defendants than African American defendants, even when accounting for the type of crime (Mustard 2001). If race is playing a role in sentencing, one might

expect that allowing for more discretion in departures would increase disparities. The data here are mixed, with one study of state courts showing greater racial disparities in states with less restrictive sentencing systems (Wang et al. 2013). Studies of the federal system, however, show no such increase after the guidelines became less restrictive in 2005 (Fischman & Schanzenbach 2012).

Another disturbing study on sentencing in Georgia demonstrated that skin color influenced judges' sentencing decisions (Burch 2015). For decades, Georgia recorded the skin tone of all criminal defendants using nine categories. As with most studies of race and sentencing, even controlling for case category and defendants' criminal history, race influenced sentences. Skin tone, however, predicted sentence length even more strongly, with light-skinned African Americans receiving shorter sentences than dark-skinned African Americans. Afrocentric features have similar consequences (King & Johnson 2016).

Experimental methods provide a mix of evidence on the influence of race on judges. Judges, like most adults, harbor implicit racial biases in that they more closely associate African Americans with negative concepts than whites (Rachlinski et al. 2009). The same study, however, shows that these implicit biases only sometimes correlate with judgment. Notably, when the race of the defendants was clearly identified, white judges treated African American and white defendants identically. When the study only suggested the race of the two defendants, however, disparities based on implicit attitudes emerged.

In one experiment, judges' willingness to convict a defendant of battery was significantly influenced by the interaction between the judge's and defendant's race. White judges convicted white and black defendants at relatively equal rates (73% to 80%), but black judges were much more likely to convict white defendants than black defendants (92% to 50%) (Rachlinski et al. 2009). In a similar problem, white adults were more likely to convict black defendants than white defendants (90% to 70%) (Sommers & Ellsworth 2001). Even where outcomes are similar, reasoning may vary by judicial race. Although white and black judges struck down the federal sentencing guidelines, black judges were significantly more likely to adopt a due process theory that white judges overwhelmingly rejected (Sisk et al. 1998).

Less is known about racial disparities outside the criminal setting. One large study showed that African American plaintiffs fare worse than white plaintiffs (Chew & Kelley 2013). Furthermore, this effect was smaller among African American judges. One experimental study of bankruptcy judges also found no effect of race (Rachlinski et al. 2006).

Summary

Overall, the body of research on the potential for invidious biases in judges arising from reliance on emotion or implicit stereotypes supports a troubling conclusion: Judges do not easily set such extralegal matters aside. The feelings and biases that influence most adults seem to also affect judges.

CONCLUSION

Society expects a lot from its judges. Most people have opinions and beliefs that arise from their identities, but we expect judges to set these aside. Most people overreact to mechanisms of accountability, but we expect judges to respond in a measured fashion. Most people rely too heavily on intuition and heuristics, but we expect judges to be deliberative and logical. Emotions and biases influence people too readily, but we expect judges to decide within the law and on the record. The research suggests that judges fall short in some respects.

Can judges do more to avoid these problems? At the moment, most of the relevant research emphasizes judicial departures from the expectation that judges will be wholly impartial. Research on decision-making strategies in general, however, provides some guidance on improving the quality of judicial decision making. Most of the deviations from objectivity that the research uncovers can be traced to an excessive reliance on intuitive reasoning. Judges relying on snap judgments are apt to fall back on their attitudes, implicit stereotypes, emotional responses, moods, and cognitive shortcuts. Using strategies that promote more deliberative, system 2 reasoning can allow judges to avoid many of the pitfalls that the research documents in judges.

As an initial matter, judges facing extreme time pressures are apt to perform worse than judges who have enough time. Judges everywhere face crowded dockets and enormous time pressures. Justice delayed is justice denied, and the busy judge cannot always have the luxury of constantly revisiting their approach to each and every case. And many judges are rightly proud of their ability to manage their crowded dockets with dispatch. That expedience comes at a price, however. Time pressure tends to produce worse judgment in the kinds of settings in which judges operate. People under time pressure rely more on simple cognitive strategies, emotional reactions, and stereotypes (Gilbert 2002). One strategy for improving decision making is to ensure that judges do not face such overwhelming caseloads that they are pressured to make one snap judgment after another.

Judges also need more feedback. The courtroom provides little feedback to judges. Appeal is uncommon, and absent an extreme ruling, most decisions judges make cannot easily be categorized as correct or incorrect. Judges can, however, develop feedback mechanisms. Judges can keep track of their sentences, outcomes of their bail hearings, and the like to identify sources of inconsistency or bias. Research on umpires in Major League Baseball, for example, demonstrates that home-plate umpires who know their calls are being recorded and scored by machine do not express racial preferences in calling balls and strikes (in contrast to unmonitored umpires) (Parsons et al. 2011). Judges would be well advised to create such information for themselves, as major newspaper organizations today are obtaining and reporting it.

Judges can adopt some simple strategies to use more deliberative cognitive processes in the courtroom. Writing orders and opinions is one such strategy (Oldfather 2008). Most judges have had the experience of having an idea of how to decide a case but then having trouble drafting an opinion to support that conclusion. That conflict reflects a divergence between intuitive and deliberative reasoning and suggests that the judge's initial impression might be wrong. Using the writing process to facilitate deliberative reasoning need not be time consuming. Even drafting a brief account of the reasoning that is not shared with the litigants would suffice to trigger a deliberative approach.

Other mechanisms to trigger slower, more careful reasoning are available to even the busy judge. Judges can draft scripts or checklists to follow in matters that commonly come before them, which facilitates a more thorough reasoning process. Asking the parties to argue each side of the matter carefully, even without written briefs, can force the judge to compare both sides of a dispute in deliberative fashion. Judges should also be mindful of the need to take frequent breaks, so as to avoid fatigue. To be sure, all of these methods add some time to the decision-making process a judge must follow, but not so much as to bog down a busy courtroom.

Overall, the research on judges supports Jerome Frank's (1949, p. 410) assertion some 70 years ago: "When all is said and done, we must face the fact that judges are human." Frank was not at all pessimistic about this fact, and neither are we. Understanding judges' weaknesses is not a means of mocking them, nor should it undermine faith in the system of justice. Rather, it is a means of understanding the circumstances under which judges are vulnerable and furthering the never-ending struggle toward a fair and impartial system.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

LITERATURE CITED

- Abrams DS, Bertrand M, Mullainathan S. 2012. Do judges vary in their treatment of race? *J. Legal Stud.* 41(2):347–83
- Albonetti CA. 1997. Sentencing under the federal sentencing guidelines: effects of defendant characteristics, guilty pleas, and departures on sentence outcomes for drug offenses. *Law Soc. Rev.* 31(4):789–822
- Allen DW, Wall DE. 1987. The behavior of women state supreme court justices: Are they tokens or outsiders? *Justice Syst. J.* 12(2):232–44
- Am. Bar Assoc. 2011. *Model Code of Judicial Conduct*. Washington, DC: Am. Bar Assoc.
- Am. Law Inst. 2009. *Restatement (Third) of Torts*. Philadelphia: Am. Law Inst.
- Anderson JC, Lowe DJ, Reckers PMJ. 1993. Evaluation of auditor decisions: hindsight bias effects and the expectation gap. *J. Econ. Psychol.* 14(4):711–37
- Ariely D. 2009. *Predictably Irrational: The Hidden Forces that Shape Our Decisions*. New York: HarperCollins
- Aspin LT. 2007. Retention elections and evaluations: A response to current trends in contested judicial elections? In *Future Trends in State Courts*, ed. C Flango, pp. 104–9. Washington, DC: Natl. Cent. State Courts
- Ayres I, Waldfogel J. 1994. A market test for race discrimination in bail setting. *Stanford Law Rev.* 46(5):987–1047
- Baum L. 1997. *The Puzzle of Judicial Behavior*. Ann Arbor: Univ. Mich. Press
- Beattay R, Matsuura T, Jeglic EL. 2014. Judicial bond setting behavior: The perceived nature of the crime may matter more than how serious it is. *Psychol. Public Policy Law* 20(4):411–20
- Beebe B. 2006. An empirical study of the multifactor tests for trademark infringement. *Calif. Law Rev.* 94(6):1581–654
- Beebe B. 2008. An empirical study of U.S. copyright fair use opinions, 1978–2005. *Univ. Pa. Law Rev.* 156(3):549–624
- Blanc PD, Rosenthal R, Cordell LH. 1985. The appearance of justice: judges' verbal and nonverbal behavior in criminal jury trials. *Stanford Law Rev.* 38(1):89–164
- Block BP. 1991. Just convictions or just convictions? *Issues Criminol. Legal Psychol.* 1(17):20–24
- Blume JH, Eisenberg T. 1999. Judicial politics, death penalty appeals, and case selection: an empirical study. *South. Calif. Law Rev.* 72(2/3):465–503
- Booker v. United States*, 543 U.S. 220 (2005)
- Bornstein BH, Miller ML. 2009. Does a judge's religion influence decision making? *Court Rev.* 45(3):112–15
- Boyd CL. 2013. She'll settle it? *J. Law Courts* 1(2):193–219
- Boyd CL. 2016. Representation on the courts? The effects of trial judges' sex and race. *Political Res. Q.* 69(49):788–99
- Boyd CL, Epstein L, Martin AD. 2010. Untangling the causal effects of sex on judging. *Am. J. Political Sci.* 54(2):389–411
- Boyd CL, Spriggs JF. 2009. An examination of strategic anticipation of appellate court preferences by federal district court judges. *Wash. Univ. J. Law Policy* 29(1):37–80
- Brace P, Hall MG, Langer L. 1999. Judicial choice and the politics of abortion: institutions, context, and the autonomy of courts. *Albany Law Rev.* 62(2):1265–303
- Burch T. 2015. Skin color and the criminal justice system: beyond black-white disparities in criminal sentencing. *J. Empir. Legal Stud.* 12(3):395–420
- Bushway SD, Owens EG, Piehl AM. 2012. Sentencing guidelines and judicial discretion: quasi-experimental evidence from human calculation errors. *J. Empir. Legal Stud.* 9(2):291–319
- Byrne v. Boadle*, (Exch. 1863) 2 H. & C. 722, 159 Eng. Rep. 299

- Cameron C, Cummings C. 2003. *Diversity and judicial decision making: evidence from affirmative action cases in the federal courts of appeals, 1971–1999*. Presented at Crafting Oper. Inst. Conf., April 11–13, Yale Univ., New Haven, CT
- Cann DM. 2007. Justice for sale? Campaign contributions and judicial decisionmaking. *State Politics Policy Q.* 7(3):281–97
- Caperton v. A.T. Massey Coal Co.*, 556 U.S. 868 (2009)
- Chen D, Moskowitz TJ, Shue K. 2016. Decision-making under the gambler's fallacy: evidence from asylum judges, loan officers, and baseball umpires. *Q. J. Econ.* 131(3):1181–242
- Chew PK, Kelley RE. 2009. Myth of the color-blind judge: an empirical analysis of racial harassment cases. *Wash. Univ. Law Rev.* 86(5):1117–66
- Chew PK, Kelley RE. 2013. The realism of race in judicial decision making: an empirical analysis of plaintiffs' race and judges' race. *Harvard J. Racial Ethn. Justice* 28(1):91–113
- Cohen MA. 1991. Explaining judicial behavior or what's "unconstitutional" about the sentencing commission. *J. Law Econ. Organ.* 7(1):183–99
- Cox AB, Miles TJ. 2008. Judging the Voting Rights Act. *Columbia Law Rev.* 108(1):1–54
- Daftary-Kapur T, Penrod SD, O'Connor M, Wallace B. 2014. Examining pretrial publicity in a shadow jury paradigm: issues of slant, quantity, persistence and generalizability. *Law Hum. Behav.* 38(5):462–77
- Danziger S, Levav J, Avnaim-Pesso L. 2011. Extraneous factors in judicial decisions. *PNAS* 108(17):6889–92
- Davis S. 1986. The impact of President Carter's judicial selection reforms: a voting analysis of the United States Courts of Appeals. *Am. Politics Q.* 14(3):20–44
- Dhami MK. 2003. Psychological models of professional decision-making. *Psychol. Sci.* 14(2):175–80
- Downs C, Lyons P. 1991. Natural observations of the links between attractiveness and initial legal judgments. *Personal. Soc. Psychol. Bull.* 17(5):541–47
- Ebbesen EB, Konečni VJ. 1975. Decision making and information integration in the courts: the setting of bail. *J. Personal. Soc. Psychol.* 32(5):805–21
- Ebbesen EB, Konečni VJ. 1982. An analysis of the sentencing system. In *The Criminal Justice System: A Social-Psychological Analysis*, ed. VJ Konečni, EB Ebbesen, pp. 293–332. San Francisco: W.H. Freeman
- Eisenberg T, Miller GP. 2004. Attorney fees in class action settlements: an empirical study. *J. Empir. Legal Stud.* 1(1):27–78
- Eisenberg T, Rachlinski JJ, Wells MT. 2002. Reconciling experimental incoherence with real-world coherence in punitive damages. *Stanford Law Rev.* 54(6):1239–71
- Englich B, Mussweiler T, Strack F. 2006. Playing dice with criminal sentences: the influence of irrelevant anchors on experts' judicial decision making. *Personal. Soc. Psychol. Bull.* 32(2):188–200
- Epstein L. 2016. Some thoughts on the study of judicial behavior. *William Mary Law Rev.* 57(6):2017–73
- Epstein L, Knight J. 1997. *The Choices Justices Make*. Washington, DC: CQ
- Epstein L, Knight J. 2013. Reconsidering judicial preferences. *Annu. Rev. Political Sci.* 16(1):11–31
- Epstein L, Landes WM, Posner RA. 2013. *The Behavior of Federal Judges*. Cambridge, MA: Harvard Univ. Press
- Epstein L, Martin AD, Martin AW. 2004. Does age (really) matter? A response to Manning, Carroll, and Carp. *Soc. Sci. Q.* 85(1):19–30
- Espinosa EM, Sorensen JR. 2016. The influence of gender and traumatic experiences on length of time served in juvenile justice settings. *Crim. Justice Behav.* 43(2):187–203
- Etienne M. 2010. Sentencing women: reassessing the claims of disparity. *J. Gend. Race Justice* 14(1):73–84
- Everett RS, Nienstedt BC. 1999. Race, remorse and sentence reduction: Is saying you're sorry enough? *Justice Q.* 16(1):99–122
- Farhang S, Wawro G. 2004. Institutional dynamics on the U.S. Court of Appeals: minority representation under panel decision making. *J. Law Econ. Organ.* 20(2):299–330
- Fariña F, Arce R, Novo M. 2003. Anchoring in judicial decision-making. *Psychol. Spain* 7(1):56–65
- Fischman J. 2015. Interpreting circuit court voting patterns: a social interactions framework. *J. Law Econ. Organ.* 31(4):808–42
- Fischman JB, Schanzenbach MM. 2012. Racial disparities under the federal sentencing guidelines: the role of judicial discretion and mandatory minimums. *J. Empir. Legal Stud.* 9(4):729–64

- Fox R, Van Sickel R. 2000. Gender dynamics and judicial behavior in trial courts: an exploratory study. *Justice Syst. J.* 21(3):261–80
- Frank J. 1930. *Law and the Modern Mind*. Paris: Brentano's
- Frank J. 1949. *Courts on Trial: Myth and Reality in American Justice*. Princeton, NJ: Princeton Univ. Press
- Frederick S. 2005. Cognitive reflection and decision making. *J. Econ. Perspect.* 19(4):25–42
- Gigerenzer G, Todd P. 1999. Fast and frugal heuristics. In *Simple Heuristics that Make Us Smart*, ed. G Gigerenzer, P Todd, ABC Res. Group, pp. 1–17. Oxford: Oxford Univ. Press
- Gilbert DT. 2002. Inferential correction. In *Heuristics and Biases: The Psychology of Intuitive Judgment*, ed. T Gilovich, D Griffin, D Kahneman, pp. 167–84. Cambridge, UK: Cambridge Univ. Press
- Glynn A, Sen M. 2015. Identifying judicial empathy: Does having daughters cause judges to rule for women's issues? *Am. J. Political Sci.* 59(1):37–54
- Gottschall J. 1983. Carter's judicial appointments: the influence of affirmative action and merit selection in voting on the US Court of Appeals. *Judicature* 67:165–73
- Gruhl J, Spohn C, Welch S. 1981. Women as policy makers: the case of trial judges. *Am. J. Political Sci.* 25(2):308–22
- Guthrie C, George TE. 2005. The futility of appeal: disciplinary insights into the “affirmance effect” on the United States Courts of Appeals. *Fla. State Law Rev.* 32(2):357–86
- Guthrie C, Rachlinski JJ, Wistrich AJ. 2001. Inside the judicial mind. *Cornell Law Rev.* 86(4):777–830
- Guthrie C, Rachlinski JJ, Wistrich AJ. 2007. Blinking on the bench: how judges decide cases. *Cornell Law Rev.* 93(1):1–43
- Guthrie C, Rachlinski JJ, Wistrich AJ. 2009. The “hidden judiciary”: an empirical examination of executive branch justice. *Duke Law J.* 58(7):1477–530
- Helland E, Tabarrock A. 2002. The effect of electoral institutions on tort awards. *Am. Law Econ. Rev.* 4(2):341–70
- Helm RK, Wistrich AJ, Rachlinski JJ. 2016. Are arbitrators human? *J. Empir. Legal Stud.* 13(4):666–92
- Holmes MD, Hosch HM, Daudistel HC, Perez DA, Graves JB. 1993. Judges' ethnicity and minority sentencing: evidence concerning Hispanics. *Soc. Sci. Q.* 74(3):496–506
- Huber GA, Gordon SC. 2004. Accountability and coercion: Is justice blind when it runs for office? *Am. J. Political Sci.* 48(2):247–63
- Idelman SC. 2005. The concealment of religious values in judicial decisionmaking. *Va. Law Rev.* 91(2):515–34
- Jamieson KH, Hennessy M. 2007. Public understanding of and support for the courts: survey results. *Georgetown Law J.* 95(4):899–902
- Jones C, Rankin MB. 2015. Justice as a rounding error? Evidence of subconscious bias in second-degree murder sentences in Canada. *Osgoode Hall Law J.* 52(1):109–40
- Kahneman D. 2011. *Thinking Fast and Slow*. New York: Farrar, Straus & Giroux
- Kahneman D, Tversky A. 1973. On the psychology of prediction. *Psychol. Rev.* 80(4):237–51
- Kang MS, Shepherd JM. 2015. Partisanship in state supreme courts: the empirical relationship between party campaign contributions and judicial decision making. *J. Legal Stud.* 44(S1):S161–85
- Kastellec JP. 2013. Racial diversity and judicial influence on appellate courts. *Am. J. Political Sci.* 57(1):167–83
- Kaye DH. 1979. Probability theory meets res ipsa loquitur. *Mich. Law Rev.* 77(6):1456–84
- King RD, Johnson BD. 2016. A punishing look: skin tone and Afrocentric features in the halls of justice. *Am. J. Sociol.* 122(1):90–124
- Klein DE, Hume RJ. 2003. Fear of reversal as an explanation of lower court compliance. *Law Soc. Rev.* 37(3):579–81
- Kritzer HM, Uhlman TM. 1977. Sisterhood in the courtroom: sex of judge and defendant in criminal case disposition. *Soc. Sci. Q.* 14(1):77–88
- Kunin CC, Ebbeson EB, Konecni VJ. 1992. An archival study of decision-making in child custody disputes. *J. Clin. Psychol.* 48(4):564–73
- Landsman S, Rakos RF. 1994. A preliminary inquiry into the effect of potentially biasing information on judges and jurors in civil litigation. *Behav. Sci. Law* 12(2):113–26
- Lappi-Seppälä T. 2001. Sentencing and punishment in Finland: the decline of the repressive ideal. In *Sentencing and Sanctions in Western Countries*, ed. M Tonry, R Frase, pp. 92–150. Oxford: Oxford Univ. Press

- Leibovitch A. 2016. Relative judgments. *J. Legal Stud.* 45(2):281–330
- Leifer D, Sample LL. 2010. Do judges follow sentencing recommendations, or do recommendations simply reflect what they want to hear? An examination of one state court. *J. Crime Justice* 33(2):127–51
- Lemley MA, Li S, Urban JM. 2014. Does familiarity breed contempt among judges deciding patent cases? *Stanford Law Rev.* 66(5):1121–58
- Levy I. 2013. Lightening the overload of CPR Rule 3.9. *Civil Justice Q.* 32(2):139–52
- Liberato L, Rutter K. 2003. Reasons for reversal in the Texas Courts of Appeals. *South Tex. Law Rev.* 44:431–81
- Lim CS. 2015. Media influence on courts: evidence from civil case adjudication. *Am. Law Econ. Rev.* 17(1):87–126
- Lim CS, Snyder JM Jr., Stromberg D. 2015. The judge, the politician, and the press: newspaper coverage and criminal sentencing across electoral systems. *Am. Econ. J. Appl. Econ.* 7(4):103–35
- Maccoby EE, Mnookin RH. 1992. *Dividing the Child: Social & Legal Dilemmas of Custody*. Cambridge, MA: Harvard Univ. Press
- MacDonald JM, Chesney-Lind M. 2001. Gender bias and juvenile justice revisited: a multiyear analysis. *Crime Delinq.* 47(2):173–95
- Maguire JM, Epstein CSS. 1927. Rules of evidence in preliminary controversies as to admissibility. *Yale Law J.* 36(8):1101–25
- Manning KL, Carroll BA, Carp RA. 2004. Does age matter? Judicial decision making in age discrimination cases. *Soc. Sci. Q.* 85(1):1–18
- Maroney TA. 2011. The persistent cultural script of judicial dispassion. *Calif. Law Rev.* 99(2):629–82
- Maveety N, ed. 2003. *The Pioneers of Judicial Behavior*. Ann Arbor: Univ. Mich. Press
- McCall M. 2003. Gender, judicial dissent, and issue salience: the voting behavior of state supreme court justices in sexual harassment cases, 1980–1998. *Soc. Sci. J.* 40(1):79–97
- Mitchell O. 2005. A meta-analysis of race and sentencing research: explaining the inconsistencies. *J. Quant. Criminol.* 21(4):439–66
- Moran G, Cutler BL. 1991. The prejudicial impact of pre-trial publicity. *J. Appl. Soc. Psychol.* 21(5):345–67
- Mustard DB. 2001. Racial, ethnic, and gender disparities in sentencing: evidence from the U.S. Federal Courts. *J. Law Econ.* 44(1):285–314
- Nagel SS. 1962. Judicial backgrounds and criminal cases. *J. Crim. Law Criminol.* 53(3):333–39
- Oldfather CM. 2008. Writing, cognition, and the nature of the judicial function. *Georgetown Law J.* 96(4):1283–345
- Otto AL, Penrod SD, Dexter HR. 1994. The biasing impact of pre-trial publicity on juror judgment. *Law Hum. Behav.* 18(4):453–69
- Paradise J. 2012. The disparity between men and women in custody disputes: Is joint custody the answer to everyone's problems? *St. John's Law Rev.* 72(2):517–80
- Parsons CA, Sulaeman J, Yates MC, Hamermesh DS. 2011. Strike three: umpires' demand for discrimination. *Am. Econ. Rev.* 101(4):1410–35
- Perisic JL. 2005. Female judges matter: gender and collegial decisionmaking in the Federal Appellate Courts. *Yale Law J.* 114(7):1759–90
- Pinello DR. 1999. Linking party to judicial ideology in American courts: a meta-analysis. *Justice Syst. J.* 20(3):219–54
- Pinello DR. 2003. *Gay Rights and American Law*. Cambridge, UK: Cambridge Univ. Press
- Posner RA. 2008. *How Judges Think*. Cambridge, MA: Harvard Univ. Press
- Priest GL, Klein B. 1984. The selection of disputes for litigation. *J. Legal Stud.* 13(1):1–55
- Rachlinski JJ, Guthrie C, Wistrich AJ. 2006. Inside the bankruptcy judge's mind. *Boston Univ. Law Rev.* 86(5):1227–65
- Rachlinski JJ, Guthrie C, Wistrich AJ. 2013a. Contrition in the courtroom: Do apologies affect adjudication? *Cornell Law Rev.* 98(5):1189–244
- Rachlinski JJ, Johnson SL, Wistrich AJ, Guthrie C. 2009. Does unconscious bias affect trial judges? *Notre Dame Law Rev.* 84(3):1195–246
- Rachlinski JJ, Wistrich AJ, Guthrie C. 2011. Probability, probable cause, and hindsight. *J. Empir. Legal Stud.* 8(S1):72–98

- Rachlinski JJ, Wistrich AJ, Guthrie C. 2013b. Altering attention in adjudication. *UCLA Law Rev.* 60(6):1586–618
- Rachlinski JJ, Wistrich AJ, Guthrie C. 2015. Can judges make reliable numeric judgments? Distorted damages and skewed sentences. *Indiana Law J.* 90(2):695–739
- Ramji-Nogales J, Schoenholtz A, Schrag PG. 2007. Refugee roulette: disparities in asylum adjudication. *Stanford Law Rev.* 60(2):295–412
- Randazzo KA. 2008. Strategic anticipation and the hierarchy of justice in the U.S. District Courts. *Am. Political Res.* 36(5):669–93
- Rehaag S. 2012. Judicial review of refugee determinations: The luck of the draw? *Queen's Law J.* 38(1):1–58
- Robinson R. 2011. Does prosecutorial experience “balance out” a judge’s liberal tendencies? *Justice Syst. J.* 32(2):143–66
- Rosenblatt A, Greenberg J, Solomon S, Pyszczyński T, Lyon D. 1989. Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. *J. Personal. Soc. Psychol.* 57(4):681–90
- Ruback RB, Vardaman PJ. 1997. Decision making in delinquency cases: the role of race and juveniles’ admission/denial of the crime. *Law Hum. Behav.* 21(1):47–69
- Ryo E. 2016. Pretrial: a study of immigration bond hearings. *Law Soc. Rev.* 50(1):117–53
- Schanzenbach M. 2005. Racial and sex disparities in prison sentences: the effect of district-level judicial demographics. *J. Legal Stud.* 34(1):57–92
- Schanzenbach MM, Tiller EH. 2006. Strategic judging under the U.S. sentencing guidelines: positive political theory and evidence. *J. Law Econ. Organ.* 23(1):24–56
- Scott RW. 2011. Race disparity under advisory guidelines: dueling assessments and potential responses. *Criminol. Public Policy* 10(4):1129–38
- Segal JA, Spaeth HJ. 2015. *The Supreme Court and the Attitudinal Model Revisited*. Cambridge, MA: Cambridge Univ. Press
- Senate Comm. Judic. 2010. *Confirmation Hearing on the Nomination of Elena Kagan to be an Associate Justice of the Supreme Court of the United States*, 111th Cong. 103 (statement of Elena Kagan, Supreme Court nomin.)
- Shepherd JM. 2009. The influence of retention politics on judges’ voting. *J. Legal Stud.* 38(1):169–206
- Shepherd JM, Kang MS. 2011. The partisan price of justice: an empirical analysis of campaign contributions and judicial decisions. *N.Y. Univ. Law Rev.* 86(1):69–130
- Sherif M, Taud D, Hovland CI. 1958. Assimilation and contrast effects of anchoring stimuli. *J. Exper. Psychol.* 55(2):150–55
- Sisk GC, Heise M. 2012. Muslims and religious liberty in the era of 9/11: empirical evidence from the federal courts. *Iowa Law Rev.* 98(1):231–91
- Sisk GC, Heise M, Morriss AP. 1998. Charting the influences on the judicial mind: an empirical study of judicial reasoning. *NYU Law Rev.* 73(5):1377–500
- Sisk GC, Heise M, Morriss AP. 2004. Searching for the soul of judicial decisionmaking: an empirical study of religious freedom decisions. *Ohio State Law J.* 65(3):491–614
- Sommers SR, Ellsworth PC. 2001. White juror bias: an investigation of prejudice against black defendants in the American courtroom. *Psychol. Public Policy Law* 7(1):201–29
- Songer DR, Davis S, Haire S. 1994. A reappraisal of diversification in the federal courts: gender effects in the courts of appeals. *J. Politics* 56(2):425–39
- Songer DR, Ginn MR, Sarver T. 2003. Do judges follow the law when there is no fear of reversal? *Justice Syst. J.* 24(3):137–61
- Songer DR, Tabrizi SJ. 1999. The religious right in court: the decision making of Christian Evangelicals in state supreme courts. *J. Politics* 61(2):507–26
- Sorauf FJ. 1976. *The Wall of Separation: The Constitutional Politics of Church and State*. Princeton, NJ: Princeton Univ. Press
- Spamann H, Klöhn L. 2016. Justice is less blind, and less legalistic, than we thought: evidence from an experiment with real judges. *J. Legal Stud.* 45(2):255–80
- Spohn C. 1990. Decision making in sexual assault cases: Do black and female judges make a difference? *Women Crim. Justice* 2(1):83–105

- Spohn C. 2013. The effects of the offender's race, ethnicity, and sex on federal sentencing outcomes in the guidelines era. *Law Contemp. Probl.* 76(1):75–104
- Spohn C, Gruhl J, Welch S. 1987. The impact of the ethnicity and gender of defendants on the decision to reject or dismiss felony charges. *Criminology* 25(1):175–92
- Starr S. 2012. Estimating gender disparities federal criminal cases. *Am. Law Econ. Rev.* 17(1):127–59
- Stebblay N, Hosch H, Culhane S, McWethy A. 2006. The impact on juror verdicts of judicial instruction to disregard inadmissible evidence: a meta-analysis. *Law Hum. Behav.* 30(4):469–92
- Steffensmeier D, Britt CL. 2001. Judges' race and judicial decision making: Do black judges sentence differently? *Soc. Sci. Q.* 82(4):749–64
- Steffensmeier D, Hebert C. 1999. Men and women policy makers: Does the judge's gender affect the sentencing of criminal defendants? *Soc. Forces* 77:1163–96
- Stewart JE. 1980. Defendant's attractiveness as a factor in the outcome of criminal trials: an observational study. *J. Appl. Soc. Psychol.* 10(3):348–61
- Stewart JE. 1985. Appearance and punishment: the attraction-leniency effect in the courtroom. *J. Soc. Psychol.* 125(3):373–78
- Strauss M. 1996. Sequestration. *Am. J. Crim. Law* 24(1):63–122
- Sunstein CR, Schkade D, Ellman LM, Sawicki A. 2006. *Are Judges Political? An Empirical Analysis of the Federal Judiciary*. Washington, DC: Brookings Inst. Press
- Tate CN, Handberg R. 1991. Time binding and theory building in personal attribute models of Supreme Court voting behavior, 1916–88. *Am. J. Political Sci.* 35(2):460–80
- Tversky A, Kahneman D. 1974. Judgment under uncertainty: heuristics and biases. *Science* 185(4157):1124–31
- United States v. Ballard*, 322 U.S. 78, 93–94 (1944) (J Jackson, dissenting)
- Viscusi WK. 1999. How do judges think about risk? *Am. Law Econ. Rev.* 1(1):26–62
- Walker TG, Barrow DJ. 1985. The diversification of the federal bench: policy and process ramifications. *J. Politics* 47(2):596–617
- Wallsten TS. 1981. Physician and medical student bias in evaluating diagnostic information. *Med. Decis. Mak.* 1(1):145–64
- Wang X, Mears DP, Spohn C, Dari L. 2013. Assessing the differential effects of race and ethnicity on sentence outcomes under different sentencing systems. *Crime Delinq.* 59(1):87–114
- Wason PC. 1968. Reasoning about a rule. *Q. J. Exp. Psychol.* 20(3):273–81
- Weber EU, Johnson EJ. 2009. Mindful judgment and decision making. *Annu. Rev. Psychol.* 60(1):53–85
- Weinberg JD, Nielsen LB. 2011. Examining empathy: discrimination, experience, and judicial decisionmaking. *South. Calif. Law Rev.* 85(2):313–52
- Weinshall-Margel K, Shapard J. 2011. Overlooked factors in the analysis of parole decisions. *PNAS* 108(42):E833
- Weisselberg CD, Dunworth T. 1993. Inter-district variation under the Guidelines: The trees may be more significant than the forest. *Fed. Sentencing Rep.* 6(1):25–28
- Wissler RJ, Hart AJ, Saks MJ. 1999. Decisionmaking about general damages: a comparison of jurors, judges, and lawyers. *Mich. Law Rev.* 71(3):751–821
- Wistrich AJ. 2010. Defining good judging. In *The Psychology of Judicial Decision Making*, ed. D Klein, G Mitchell, pp. 249–67. Oxford: Oxford Univ. Press
- Wistrich AJ, Guthrie C, Rachlinski JJ. 2005. Can judges ignore inadmissible information? The difficulty of deliberately disregarding. *Univ. Pa. Law Rev.* 153(4):1251–385
- Wistrich AJ, Rachlinski JJ. 2013. How lawyers' intuitions prolong litigation. *South. Calif. Law Rev.* 86(3):571–603
- Wistrich AJ, Rachlinski JJ, Guthrie C. 2015. Heart versus head: Do judges follow the law or follow their feelings? *Tex. Law Rev.* 93(4):855–923
- Zebrowitz LA, McDonald SM. 1991. The impact of litigants' baby-facedness and attractiveness on adjudications in small claims courts. *Law Hum. Behav.* 15(6):603–23
- Zorn C, Bowie JB. 2010. Ideological influences on decision making in the federal judicial hierarchy: an empirical assessment. *J. Politics* 72(4):1212–21