(SUB)CONSCIOUS JUDGING
Elizabeth Thornburg*

Abstract

Fact inferences made by the trial judge are the lynchpin of civil litigation. If inferences were a matter of universally held logical deductions, this would not be troubling. Inferences, however, are deeply contestable conclusions that vary from judge to judge. Subconscious psychological phenomena can lead to flawed reasoning, implicit bias, and culturally influenced perceptions. Inferences differ significantly, and they matter. Given the homogeneous makeup of the judiciary, this is a significant concern.

This article will demonstrate the ubiquity, importance, and variability of inferences by examining actual cases in which trial and appellate (or majority and dissenting) judges draw quite different inferences from the same record. It will then review the psychological literature to show ways in which judges are affected by subconscious forces. It concludes by suggesting reforms to judicial education, use of decision mechanisms that promote conscious deliberation, and civil procedure rule changes designed to increase information and decrease the impact of individual judges’ inferences.

TABLE OF CONTENTS

I. INTRODUCTION
II. DEFINING “INFECTION” AND INFECTION DRAWING
III. A TALE OF TWO JUDGES
IV. JUDICIAL INFERENCES IN REAL LIFE
   A. Dispositive Motions
   B. Other Pretrial Rulings
V. JUDGES AND HEURISTICS
   A. Introduction to Heuristics: Unconscious Shortcuts
   B. Heuristics Affect Decisions
   C. When Does Professional Expertise Intervene?

* Richard R. Lee Endowed Professor of Law, Altshuler Distinguished Teaching Professor, SMU Dedman School of Law. Many thanks for helpful comments to Professors Kevin Clermont, Deborah Hensler, Elizabeth Schneider, Brooke Coleman, Suja Thomas, Scott Dodson, Joe Glannon, Grant Hayden, Jenia Turner, the participants in the SMU Dedman faculty workshop, and my fellow panelists at the Feminist Theory and Civil Procedure Roundtable at the 2018 Law & Society meeting. Thanks also for both consultation and comments to experts at the Federal Judicial Center, including former Director Jeremy Fogel, Senior Research Associates Jason Cantone and Beth Wiggins, and Senior Judicial Education Attorney Denise Neary. Finally, thanks to my wonderful liaison librarian, Donna Wolff, for research support. This paper was supported by the Lidji Endowment Fund.
D. Implicit Bias
E. Cultural Cognition
F. Judicial Demographics
G. Summary

VI. IMPROVING THE SYSTEM
A. Judicial Education
B. Deliberation-Prompting Practices
C. Procedure Rules

VII. CONCLUSION
(SUB)CONSCIOUS JUDGING

"[T]he test for determining whether an inference (from circumstantial evidence) is a rational one is stated in terms of mathematical precision but is one which allows the very greatest latitude in actual application. . . . [T]he authoritative language of nice and scientific precision in which such conclusions are cast is after all only the language of delusive exactness."
– Fleming James, Jr.1

“We may try to see things as objectively as we please. None the less, we can never see them with any eyes except our own.”
– Benjamin Cardozo2

“Determining whether a complaint states a plausible claim for relief will, . . . be a context-specific task that requires the reviewing court to draw on its judicial experience and common sense.”
-- Justice Anthony Kennedy3

“[M]otivated cognition [is] the ubiquitous tendency of people to form perceptions, and to process factual information generally, in a manner congenial to their values and desires.”
– Dan M. Kahan4

I. INTRODUCTION

Inferences about facts are an integral and unavoidable part of civil litigation. As in life, much of what we “know” are conclusions, formed by drawing inferences from a collection of direct and circumstantial evidence. Juries, for example, are instructed that “inferences are simply deductions or conclusions which reason and common sense lead the jury to draw from the evidence received in the case.”5 At trial, then, juries perform the intertwined functions of deciding what evidence to believe, what inferences to draw (and not draw), and how the law applies to the facts it has found.6

1 Fleming James, Jr., Functions of Judge and Jury in Negligence Cases, 58 YALE L.J. 667, 673-74 (1949).
6 In doing so, they will apply the applicable standard of proof, generally “preponderance of the evidence” in civil cases. See generally Kevin M. Clermont, Staying Faithful to the
Trials, however, are not the only stage at which inferences play a role, and judges rather than juries consciously and unconsciously draw inferences that will shape the course of the litigation and the parties’ likelihood of success. This has long been recognized in the context of the pivotal pretrial rulings that take the case away from the jury: dismissals on the pleadings, and summary judgment. Less recognized are the many other ways in which judges make decisions based on facts during the pretrial period, and the ways in which they employ inferences in doing so. Although they do not technically end litigation, decisions about issues such as discovery, joinder of claims and parties, and class action status can have an enormous impact on the viability and scope of litigation. Even decisions that are labelled as exercises of discretion rather than as fact-finding are often undergirded by the judge’s belief about facts, and those beliefs are formed after drawing inferences.

This article will illuminate and critique the power of judicial inferences, and will conclude by suggesting some changes in training, process, and procedure rules that could improve the inference-drawing process. While much law review literature about inferences focuses on lawmaking, this article will focus on fact finding; while much focuses on appeals, this article will look at the pretrial stage. Trial judges’ myriad decisions are riddled with inferences, and their role in finding facts makes them the most important actors in influencing the course and outcome of lawsuits.7 Section I briefly defines “inference” as it is used in this article and discusses the inference-drawing process. It highlights the powerful role of each judge’s own experience and world view in that process.

The next two sections document judges drawing pretrial inferences and show why that matters. Section III uses a hypothetical case to illustrate the types of pretrial decisions in which the trial judge’s choices about inferences can be decisive. Section IV takes the discussion from the theoretical into real courts and decisions. And both sections demonstrate that judges may find different facts because they have chosen different inferences, even though dealing with the same records.

7 This article will focus on the federal courts and federal procedure, but the same arguments would apply to state courts. Indeed, since many state court judges are faced with much larger caseloads and far fewer resources, it is possible that the need for fast decisions and tendency toward oral rather than written opinions would exacerbate the problems discussed in this article.
After these sections establish the pervasiveness of inferences and of the potential impact of disagreements, Section V brings together judicial decisions and psychological research. It begins by introducing the psychology\textsuperscript{8} and behavioral economics\textsuperscript{9} scholarship that describe the ways in which the human brain processes information and reaches decisions, including the use of heuristics.\textsuperscript{10} It then demonstrates that judges, like all other humans, are influenced in their thinking by factors such as heuristics, implicit biases, and cultural cognition. Nor does their combination of legal education, judicial experience, and judicial education overcome those effects in most cases.\textsuperscript{11} Further, empirical studies linking judicial decisions with the judges’ own race and gender provide evidence that inferences may be especially problematic in cases where race, gender, and other characteristics are salient. Given the homogeneous makeup of the federal judiciary,\textsuperscript{12} this is a significant concern.

Section VI suggests some reforms on the individual and systemic levels that might improve judicial inferences: 1) provide judicial education designed to raise judges’ awareness of their own inference-making and to increase their knowledge of the experiences and worldviews of others; 2) implement mechanisms that activate and facilitate deliberative mental processes; and 3) change the interpretation and application of the procedure rules to support rather than reject alternative inferences and decrease the impact of a single judge’s intuitions. Section VII concludes with overall lessons on recognizing and offsetting the inevitable impact of cognitive biases on fact finding.

II. DEFINING “INFEERENCE” AND INFEERENCE DRAWING

The kind of inferences and inference-drawing discussed in this article

\textsuperscript{8} See generally DANIEL KAHNEMAN, THINKING, FAST AND SLOW (2011).
\textsuperscript{9} See generally CASS R. SUNSTEIN, BEHAVIORAL LAW AND ECONOMICS (2000).
\textsuperscript{10} A “heuristic” is “a mental shortcut that allows people to solve problems and make judgments quickly and efficiently. These rule-of-thumb strategies shorten decision-making time and allow people to function without constantly stopping to think about their next course of action. Heuristics are helpful in many situations, but they can also lead to cognitive biases.” Kendra Cherry, Heuristics and Cognitive Biases, https://www.verywellmind.com/what-is-a-heuristic-2795235 (Nov. 13, 2018).
\textsuperscript{11} Not only is the common cliché that judges just “call balls and strikes” inaccurate, some studies show that NBA referees and MLB umpires show implicit bias effects in their officiating. Jerry Kang et al., Implicit Bias in the Courtroom, 59 U.C.L.A. L. REV. 1124, 1178 (2012).
\textsuperscript{12} See text accompanying notes xx infra.
are what Black’s Law Dictionary refers to as “inferential fact” – a “fact established by conclusions drawn from other evidence rather than from direct testimony or evidence; a fact derived logically from other facts.”¹³ That derivation goes beyond the literal limits of the testimony. As Professor Dan Simon has pointed out, “An inference is typically defined as any cognitive process of reasoning, in which a person goes beyond some known data to generate a new proposition. The result of an inference, then, is the addition of information to the person’s mental representation of an issue.”¹⁴ For example, evidence might consist of a properly authenticated videotape of a bank robbery showing the robber wearing a yellow rain hat as well as testimony from a police officer that a search of the defendant’s home produced a yellow rain hat (introducing into evidence the hat, which looks like the one on the video). An inference from that evidence might be that it is the same hat, and, further, that the defendant robbed the bank.

What is the process by which inferences are drawn, or not drawn? Scholars describe it differently, but all recognize that inferences spring from the decisionmaker’s beliefs about the world, their “generalizations.” Rationalist evidence analysis looks at evidence and inferences as logical chains, considering each piece of evidence together with generalizations that are said to justify (or not justify) an inference.¹⁵

Cognitive psychologists describe the fact-finding process, including inferences, as based on the construction of stories.¹⁶ Their experiments show that jurors impose a narrative organization on trial evidence. The ways in which they do so require jurors to use their knowledge of the world – their


¹⁵ Terence Anderson, David Schum, & William Twining, *Analysis of Evidence* 100 (2d ed. 2005) (“Every inference is dependent upon a generalization.”). See also David H. Kaye, *What is Bayesianism? A Guide for the Perplexed*, 28 JURIMETRICS J. 161 (1988) (noting that even under a mathematical model of reasoning such as Bayesian theory, the results are affected both by people’s prior beliefs and by the likelihood ratios they assign to new pieces of information). For a helpful chart showing the interaction of evidence and generalizations (and also the impact of information relevant to credibility decisions), see Anderson at 61, Figure 2.5.

generalizations – plus their knowledge about the expected structure of stories in reacting to trial evidence.17 “Analyses of inference chains leading to story events reveal that intermediate conclusions are established by converging lines of reasoning which rely on deduction from world knowledge, analogies to experienced and hypothetical episodes, and reasoning by contradiction.”18

The judicial system’s confidence in generalizations is based on an assumption that judges and jurors share a body of knowledge that will make inferences fairly uniform. But this assumption is problematic.

This . . . is commonly described as “general experience, “background knowledge,” “common sense,” or “society’s stock of knowledge.” . . . [W]e need to ask, whose experience, sense or knowledge. . . . The bases for such generalizations are as varied as the sources for the beliefs themselves – education, direct experience, the media, gossip, fiction, fantasy, speculation, prejudice, and so on.19

Judges, who like jurors operate by using human cognition, can be expected to use the same kind of evidence-generalization links to analyze inferences, and to be similarly dependent on their own experiences in forming, using, and analyzing generalizations. When judges draw inferences, or when they decide whether a reasonable jury could draw an inference, they are injecting not just logic but also their own store of generalizations into the mix. This is doubly important because the kind of reasoning involved in the inference process often operates intuitively. “The reasoner does not consciously identify the generalizations upon which her inferences depend.”20


18 Id. at 524. “Because all jurors hear the same evidence and have the same general knowledge about the expected structure of stories, differences in story construction must arise from differences in world knowledge; that is, differences in experiences and beliefs about the social world.” Pennington & Hastie, supra note xx at 525.

19 ANDERSON, supra note xx at 269.

20 Id. at 101.
III. A TALE OF TWO JUDGES

In order to try to make the impact of fact inferences clear, consider the ways in which two different judges might react to pretrial issues raised in the same lawsuit.

Two hypothetical judges, Judge A and Judge B, will rule on various motions. The point is not that one or the other judge is correct, but that their inferences may well vary significantly, with important consequences for the course of the litigation. For both of our hypothetical judges, the lawsuit begins with a complaint making the following allegations.

Acme Motors designs, manufactures, and sells an all-electric car, called the Greeny, which debuted in 2012. The Greeny was very popular with eco-conscious buyers, and all was going well for a number of years. Unfortunately, in 2016 the older model Greeny cars began to have a problem. Drivers began to experience unintended acceleration – without warning the car accelerates, and braking will not stop it.

Plaintiff David Driver purchased a new Greeny in 2013. He believes that this defect in his 2013 Greeny is what caused a horrible crash in 2017 that left him with persistent back pain, lost past income and future earning capacity, and unpaid medical bills. Driver alleges that the unintended acceleration problem is caused by electronic failures and an inadequate fault detection system, and that a safer design would include a brake override system. His complaint asserts legal theories of negligence, product liability (defective design), and gross negligence, and he seeks both compensatory and punitive damages.

Co-Plaintiff Oscar Owner also purchased a new 2013 Greeny. He has not experienced the unintended acceleration problem or been injured, but he alleges that the value of his Greeny is dramatically lower than Acme represented due to this undisclosed defect and argues that Acme has breached its warranty and violated state consumer protection statutes by failing to disclose known problems with the car. He seeks compensation for the reduced resale value of his 2013 Greeny.

Now consider a series of pretrial decisions and the ways in which inferences play a role.

A. Motion to Dismiss
Acme has denied all of the plaintiffs’ claims, and also alleges that Driver’s accident was caused by his own negligence. In addition, Acme has filed a motion to dismiss under Rule 12(b)(6), arguing that the plaintiffs fail to state a claim because they have not identified a specific defect responsible for unintended acceleration, and thus have not alleged sufficient facts to make plaintiffs’ claims plausible.

Judge A denies the motion to dismiss, finding the allegations that the car unexpectedly accelerated and would not brake, and that an alternative braking system would have stopped the car, to be facts rather than conclusions. Based on those pleaded facts, Judge A believes that inferences of negligence and defective design are at least as plausible as other inferences. In addition, Judge A finds that the allegations that prior to 2013 other Greeny owners have complained to Acme about injuries from unintended acceleration are sufficient to support a plausible inference of gross negligence. With regard to Owner’s warranty claims, Judge A is satisfied that it is reasonable to infer a causal link between the pattern of unexpected acceleration and the decrease in resale value of Owner’s car. In summary, Judge A holds that the plaintiffs’ complaint meets the requirements of Rule 8 and that Acme’s motion seeks more specificity than is required at the pleading stage.21

Judge B is concerned that the complaint does little more than recite the elements required for negligence and design defect claims and treats many of the allegations as conclusions rather than facts. Even treating them as facts, Judge B is especially unconvinced by the inference that, based on customer complaints, Acme knew or should have known of a general problem with unintended acceleration. Judge B grants the motion to dismiss. Judge B does, however, grant plaintiffs leave to amend to identify more specifically the defect and safer alternative design and the sources of Acme’s knowledge of a pattern of unintended acceleration, as well as facts tying publicity about the acceleration to changes in used Greeny prices.

B. Joinder of Parties and Claims

Assume that both judges have allowed this suit to go forward. Acme has filed a motion to sever the claims of Driver and Owner, arguing that they do not arise out of the same transaction or occurrence as required by Rule 20.

Judge A denies the motion. Judge A believes that some of the

evidence of breach, defect, and causation that will support the inferences needed for the personal injury claims will also support a strong inference with respect to the decreased value of the Greeny, and that evidence of Acme’s knowledge (part of Driver’s gross negligence claim) will also support an inference of a knowing violation of the state’s consumer law. Based on his beliefs about the validity of those inferences from common evidence, Judge A concludes that there is both a logical relationship and significant evidentiary overlap between Driver’s and Owner’s claims.

Judge B, on the other hand, grants the motion to sever. Judge B concludes that evidence about the nature of Acme’s design and knowledge of risks of harm has only a tangential relationship to the resale value of Greeny vehicles, because any inference that the nature of the defect affected the resale value is weak at best. Instead, Judge B believes that a plausible inference of loss of value would require proof of lowered resale prices immediately after the public claims about the unintended acceleration problem (whether or not such a problem really exists). In the absence of plausible inferences from shared evidence, Judge B concludes that the requirements of Rule 20 have not been met.

C. Discovery

Initial disclosures have been made, and formal discovery is now under way. For example, the plaintiffs have sent Acme a request for production of documents that includes a request for “all communications received by Acme from Greeny purchasers claiming that a Greeny car from model years 2012-15 experienced unintended acceleration.” (All parties agree that the Greeny’s design did not change in any relevant way during this time period.) Acme has objected that the request is irrelevant and asserted that the cost of compliance would be disproportionate. The parties have conferred but been unable to reach agreement, so the plaintiffs have filed a motion to compel production of the requested documents.

Judge A grants the motion to compel. Judge A sees a potential for reasonable inferences from other drivers’ experiences to plaintiffs’ claims. Judge A believes that the existence of other unintended acceleration accidents supports an inference that Driver’s car suffered from the same problem. In addition, the complaints themselves might show that Acme had been notified of a troubling pattern of malfunctions leading to injuries, which could support the inference required to prove gross negligence. Finally, thinks Judge A, the pattern of complaints might support an inference of the extent of the impact of the alleged defect on resale value. With regard to proportionality, because
of Judge A’s assessment of the strength of the inference from other complaints to the elements of plaintiffs’ claims, the information sought is very important to resolving the issues. Judge A also infers from these two plaintiffs’ claims and the nature of the national car market that the Greeny’s safety issues may be causing nationwide problems, making the issues at stake important and the overall amount in controversy quite large. Judge A believes from general experience that individual plaintiffs likely have fewer resources and less access to information than does Acme and that Acme has likely overstated the cost of compliance with this discovery request. All in all, Judge A concludes that the information sought is relevant and proportional to the needs of the case.

Judge B’s approach is quite the opposite. Since Judge B granted the motion to sever, this version of the lawsuit includes only plaintiff Driver’s claims. Judge B believes that complaints from other Greeny owners support only a very weak inference, if any, about what happened to Driver’s car. Judge B thinks Driver will instead need to show plausible inferences from the evidence of the condition of Driver’s own car, the accident involving Driver, and evidence about Acme’s design of the 2013 vehicle. Nor does Judge B find an inference from complaint letters to gross negligence at all convincing. As to proportionality, Judge B (based on the assessment that the inferences are weak) believes that the requested complaints would not be at all important in resolving the liability issues. Although Judge B thinks a nationwide product defect could be significant, this discovery request is insufficiently tied to that, and the amount in controversy is limited to Driver’s own damages. While Acme has greater access to information, Judge B believes based on the judge’s experience as a lawyer and a judge that the plaintiffs’ lawyer is probably seeking discovery about other owners’ complaints primarily to run up Acme’s costs and to seek out other potential clients. Judge B’s conclusion: the inferences from customer complaints to elements of Driver’s claims is so weak that it is only marginally relevant, and the cost of compliance far exceeds the needs of the case. Judge B denies the motion to compel.

D. Summary Judgment

It is difficult to frame a short but rich and realistic hypothetical on summary judgment. Note, though, that analysis of inferences will be critical to the resolution of a summary judgment motion. Assume that Acme has filed a motion for partial summary judgment on Driver’s product liability claim, arguing that Driver will not have sufficient evidence to support an inference of defective design. Assume further that part of the summary judgment
dispute will include a decision about whether plaintiff’s engineering expert satisfies the requirements of *Daubert*\(^22\) and will be allowed to testify.

Even the decision about whether to allow an expert to testify requires evaluation of inferences. When the question is whether the inferences to which the expert wishes to testify meet the admissibility requirements of *Daubert*, judges as gatekeepers evaluate scientific inferences.\(^23\) The case-specific factual underpinnings of the expert opinion might also differ, depending on what information has been unearthed during discovery, and a judge might also reject the reliability of an expert’s inferences based on fewer bits of information, which in turn relates back to the discovery differences highlighted above. Judge A might ultimately rule that the expert’s testimony can be considered, while Judge B might rule that it cannot.

It would not be surprising, then, if Judge A were to deny the summary judgment motion, finding that a reasonable jury could choose to draw the inferences needed for findings of design defect from the plaintiff’s circumstantial evidence and expert testimony, and for Judge B to grant the motion, concluding that inferences of Acme’s liability are too weak in light of alternative inferences of Acme’s lack of fault and Driver’s contributory negligence.

**E. Putting it Together**

Looking at all of these pretrial decisions, one can see that inference-drawing across a series of pretrial rulings can have a tremendous influence on the course of litigation. It is clearest that judicial inferences play a decisive role in rulings on dispositive motions. However, inferences about facts also play a significant role in other procedural rulings, in ways that often lie beneath exercises of discretion and balancing tests.\(^24\)

This thought experiment, though, is an academic construct. Are inferences similarly pervasive in real cases? Section IV will use examples


\(^{24}\) Even though many such decisions would be reviewed on appeal using an “abuse of discretion” standard of review, the propriety of the exercise of discretion is intertwined with underlying fact-finding, which in turn depends on the court’s inferences. *See, e.g.*, the discussions of the discovery orders in Cazorla v. Koch Foods of Mississippi, LLC, No. 3:10cv135-DPJ-FKB, 2014 WL 12639863 (S.D. Miss. Sept. 22, 2014), vacated and remanded, 838 F.3d 540 (5th Cir. 2016) and Kuttner v. Zaruba, 819 F.3d 970 (7th Cir. 2016), discussed in the text accompanying notes xx *infra*. 
from actual litigation to highlight inferences at work in various pretrial contexts. And just as was true for the hypothetical dueling judges, real cases show judges finding different facts from the same record because they disagree about what inferences should be drawn.

IV. JUDICIAL INFERENCEs IN REAL LIFE

Real cases don’t come in identical pairs for purposes of comparing inferences, but trial court decisions involving inferences are sometimes appealed, and the contrast between the trial court and appellate court decisions can turn, even in procedural rulings, on judges’ disagreements about the plausibility of inferences. Similarly, appellate decisions with a dissent may reveal that judges on the panel have chosen different inferences, and hence different facts. This section will examine some contrasting opinions as examples.

A. Dispositive Motions

The impact of inference is at its most obvious in rulings on dispositive motions – motions to dismiss for failure to state a claim under Rule 12(b)(6)

25 The cases highlighted in this section of the text were identified by searching the Westlaw “Federal District Courts” database for cases decided in the prior twelve months. I searched for cases on particular topics (e.g. discovery, summary judgment, class actions, etc.), and then read all of them to find examples of cases in which inferences played an important role. Within that set, I searched for cases that had been reversed on appeal, or for cases in which there was a dissent in the court of appeals. From those results, I chose cases that would be good illustrations of the impact of inferences. This does not claim to be either a random or comprehensive study. It would be interesting to see a more quantitative study, perhaps of all the cases decided in a particular district over a given time period.

26 Professor Suja Thomas has sensibly suggested that fact-based summary judgment appeals with dissents should be treated as per se indications that a reasonable jury could find for the nonmoving party. But at present that is not the law, and a factual dispute can be considered a matter of “law” despite a disagreement among judges. See Suja Thomas, Reforming the Summary Judgment Problem: The Consensus Requirement, 86 FORDHAM L. REV. 2241 (2018). Decisions in the courts of appeal with dissents do show both inferences at play and judges who would choose different inferences from the same record. See, e.g., Tyree v. Foxx, 835 F.3d 35 (1st Cir. 2016) (majority and dissent draw different inferences about relevance of decisions similar to, but distinct from, the one that affected plaintiff); In re Ohio Execution Protocol Litig., 845 F.3d 231 (6th Cir. 2016) (majority and dissent disagree about whether circumstantial evidence supports inference that disclosure of identity of manufacturer of lethal injection drug would make it difficult for state to acquire such drugs).

27 There are a huge number of pleading cases in which inferences play a key role—not surprising since the word “inference” is baked into the legal test created by Justice Kennedy’s opinion in Iqbal. See, e.g., Johnson v. City of New York, 16-CV-6426(KAM)(VMS), 2018 WL 1597393 (E.D.N.Y. Mar. 31, 2018) (approving inferences in favor of plaintiff in
and motions for summary judgment. As with the hypothetical judges, the point is not that some judges are right and others are wrong, but that the same record can lead judges to opposite conclusions, based in large measure on the ways in which their “experience and common sense,” operating through subconscious heuristics and biases, guide their inferences. The differences can be observed on at least two levels: 1) which evidence in the record the judges notice and consider; and 2) which inferences the judges believe could reasonably be drawn from that evidence. In the pleading context, evaluation of inferences includes a comparative process; in order to keep the case alive, the nonmovant must convince the judge that the inference it needs is at least as believable as alternative inferences.

1. Pleadings Motions: Lewis v. Bentley

In ruling on motions to dismiss for failure to state a claim, the court is instructed to analyze the complaint to identify only pleaded “facts,” and then to determine whether the inferences the lawsuit needs to proceed can plausibly be drawn from those facts. As Justice Kennedy stated:

Determining whether a complaint states a plausible claim for relief will . . . be a context-specific task that requires the reviewing court to draw on its judicial experience and common sense. But where the well-pleaded facts do not permit the court to infer more than the mere possibility of misconduct, the complaint has alleged—but it has not “show[n]”—“that the pleader is entitled to relief.”

The requirement to go beyond “mere possibility” involves a comparison of inferences. Later in the opinion, Justice Kennedy rejects the plaintiffs’ inferences because there are “more likely explanations.” In ruling on 12(b)(6) motions, then, the trial court will be deciding which inference it finds plausible.
more believable (and, using a de novo standard of review, the court of appeals will be doing the same analysis).\textsuperscript{31}

\textit{Lewis v. Bentley} arose out of a dispute between the State of Alabama and the city of Birmingham over the city’s desire to adopt a minimum wage higher than the federally mandated $7.25 per hour. Birmingham passed a local ordinance with a higher minimum wage, but the State, just as Birmingham’s ordinance was set to go into effect, enacted the “Alabama Uniform Minimum Wage and Right-to-Work Act.” According to its terms, the Act was intended to “ensure that [labor] regulation and policy is applied uniformly throughout the state.” It went beyond minimum wage law to establish the state’s “complete control” over not only minimum wage policy, but also most other employment law issues.\textsuperscript{32}

A number of Birmingham residents and public interest groups filed suit, alleging that the Act had the purpose and effect of transferring control of employment in Birmingham from municipal officials elected by a majority-black local electorate to statewide legislators elected by a majority-white electorate, with the intent of discriminating against the people who live and work in Birmingham on the basis of race.\textsuperscript{33} The district court found that Birmingham’s complaint failed to plausibly allege intentional discrimination and dismissed the case on the pleadings. Focusing on the city’s equal protection claim, the Eleventh Circuit reversed.

Disagreements about inferences explain these divergent results. The trial judge – U.S. District Judge David Proctor\textsuperscript{34} – focused on the supremacy of state government over municipal government, and noted that other states had adopted similar laws. To Judge Proctor, these race-neutral explanations were more believable than an inference of intentional discrimination.\textsuperscript{35} He found that there were “obvious alternative explanations” suggesting lawful conduct.\textsuperscript{36}

Contrast that with the Eleventh Circuit’s opinion – written by Judge

\textsuperscript{32} 2017 WL 432464 at *1.
\textsuperscript{33} \textit{Id.} at *2.
\textsuperscript{34} David Proctor, Ballotpedia, https://ballotpedia.org/David_Proctor (last visited Sept. 11, 2018).
\textsuperscript{35} \textit{Id.} at *13.
\textsuperscript{36} 2017 WL 432464 at *11.
Charles Wilson\textsuperscript{37} -- which both highlighted different evidence and found different inferences to be plausible:

\begin{itemize}
  \item “Birmingham, [is] home to the largest black population in Alabama (72%), which is reflected in the racial composition of its city council.”\textsuperscript{38}
  \item The bill was introduced by a white state representative from a town where only 1.5% of the residents are black, with eventual support fifty-two additional sponsors, all of whom were white.\textsuperscript{39}
  \item No black member of the Alabama legislature voted in favor of the bill.\textsuperscript{40}
  \item The State of Alabama has a “deep and troubled history of racial discrimination,” and “has consistently impeded the efforts of its black citizens to achieve social and economic equality.”\textsuperscript{41}
  \item In terms of predictable impact, the Act “immediately denied a significant wage increase to roughly 40,000 Birmingham residents, the vast majority of whom were black.”\textsuperscript{42}
\end{itemize}

Judge Wilson and the members of the Eleventh Circuit panel evaluated inferences quite differently from Judge Proctor, holding that these “facts plausibly imply discriminatory motivations were at play.”\textsuperscript{43} They therefore found dismissal on the pleadings to have been erroneous.

2. Summary Judgment: \textit{Chadwick v. Wellpoint, Inc.}\textsuperscript{44}

Rule 56 of the Federal Rules of Civil Procedure provides that summary judgment “shall” be granted “if the movant shows that there is no genuine dispute as to any material fact and the movant is entitled to judgment as a matter of law.”\textsuperscript{45} Courts applying this test regularly recite that a dispute is “genuine” if the evidence is such that a reasonable jury could grant a verdict in favor of the nonmovant.\textsuperscript{46} While the judge should not be comparing

\begin{itemize}
  \item \textsuperscript{37}Charles Wilson, Ballotpedia, \url{https://ballotpedia.org/Charles_Wilson_(Florida)} (last visited Sept. 11, 2018).
  \item \textsuperscript{38}\textit{Lewis}, 896 F.3d at 1287-88.
  \item \textsuperscript{39}\textit{Id}. at 1288.
  \item \textsuperscript{40}\textit{Id}.
  \item \textsuperscript{41}\textit{Id}.
  \item \textsuperscript{42}\textit{Id}.
  \item \textsuperscript{43}\textit{Id}.
  \item \textsuperscript{44}Chadwick v. Wellpoint, Inc., 550 F. Supp. 2d 140 (D. Maine 2008), \textit{rev’d} 561 F.3d 38 (1st Cir. 2009).
  \item \textsuperscript{45}FED. R. CIV. P. 56(a).
  \item \textsuperscript{46}Anderson v. Liberty Lobby, Inc., 477 U.S. 242, 248 (1986).
\end{itemize}
inferences, some opinions cite the possibility of alternative inferences in rejecting the one sought by the nonmovant as unreasonable. Judicial preferences for differing inferences therefore can have a significant impact on their rulings on fact-based summary judgment motions.47

Laurie Chadwick, a longtime employee of Wellpoint, was denied a promotion. She alleged that Wellpoint “failed to promote her because of a sex-based stereotype that women who are mothers, particularly of young children, neglect their jobs in favor of their presumed childcare responsibilities,” and filed suit against the company.48 At the trial court level, the magistrate judge and district judge granted Wellpoint’s motion for summary judgment, concluding that Chadwick’s circumstantial evidence could not support a reasonable inference of discrimination based on sex.49 The First Circuit panel (using a de novo standard of review) reversed, finding that a reasonable jury could draw just such an inference.50

At the time Wellpoint chose to promote a different employee (also a woman) to the desired “Team Lead” position, Chadwick was the mother of four: an eleven-year-old child plus six-year-old triplets. Her husband was the children’s primary caregiver, and there was no suggestion in the record that


48 561 F.3d 40.

49 550 F. Supp. 2d at 142.

50 561 F.3d 40.
her work performance had ever suffered due to childcare responsibilities. Chadwick had worked in her current position at Wellpoint for seven years, and she scored 4.4 out of 5.0 on her most recent performance review.\(^{51}\)

The other finalist for the Team Leader job, Donna Oulette, also had children, ages ten and fifteen. (It is unclear, though, whether the decisionmaker was aware that Oulette was a parent.\(^{52}\) She had been in her position for a year, and she scored 3.84 on her performance review. After interviews with a panel of three employees, including Chadwick’s supervisor Nanci Miller, Oulette was chosen for the promotion.\(^{53}\)

Because the summary judgment context requires the judge to view all evidence in the light most favorable to the nonmovant, both the trial and appellate courts assumed that Chadwick had stronger qualifications than Oulette.\(^{54}\) They also recognized that society is influenced by stereotypes about working women with children. However, they disagreed decisively about the inferential power of Chadwick’s circumstantial evidence, which consisted primarily of the information narrated above plus comments made to Chadwick in the course of the decision about promotion.

First, when Miller learned that Chadwick had three 6-year-olds, she said, “Oh my—I did not know you had triplets . . . Bless you!”\(^{55}\) U.S. District Judge D. Brock Hornby\(^{56}\) considered inferences in favor of the movant – perhaps any bias reflected by this remark was a gender-neutral concern about parents rather than mothers. He looked for evidence that Miller would not have said the same thing to the father of triplets, found it lacking, and therefore rejected an inference that gender played a role in Miller’s decision. Based on his experience and his beliefs about what a jury would know, Judge Hornby inferred that “bless you” was a phrase with “ordinary meaning” – a mere “friendly exclamation” with uses ranging from a response to a sneeze

\(^{51}\) Id. at 41.

\(^{52}\) “[T]he record does not support the inference that WellPoint knew of Oulette’s status as a mother of two children, while it is uncontested that WellPoint knew of Chadwick’s children.” Id. at 42 n.4. But see the district judge’s take on this, putting the burden on the plaintiff to show a lack of awareness of Oulette’s children and, in the absence of evidence, assuming awareness. 550 F. Supp. 2d at 141 n.1.

\(^{53}\) Id. at 41 - 42.

\(^{54}\) Note, though, that the district judge conceded only that Chadwick had “somewhat better qualifications” (550 F. Supp. 2d at 142), while the appellate panel stated, “It is a fair inference that Chadwick’s qualifications significantly outweighed those of Oulette.” 561 F.3d at 42 n.3.

\(^{55}\) 550 F. Supp. 2d at 142.

to a religious invocation.  

In contrast, Judge Normal Stahl, writing for the First Circuit panel, noted that “a jury could reasonably conclude that Miller meant that she felt badly for Chadwick because her life must have been so difficult as the mother of three young children” and “that Miller’s comment suggested pity rather than respect.” This, together with additional remarks, convinced the appellate panel that a jury could reasonably infer that the promotion was denied based on an assumption that as a woman with four small children, Chadwick would not “give her all to her job.” The appellate court also found it significant that Miller learned of the children just two months before denying Chadwick the promotion.

Second, during Chadwick’s interview for the promotion, one of the interviewers made an explicit reference to Chadwick’s parenthood. In reacting to Chadwick’s answer to a question about supervising employees who failed to meet deadlines, the interviewer said, “Laurie, you are a mother would you let your kids off the hook that easy, if they made a mess in [their] room, would you clean it or hold them accountable?” District Judge Hornby again saw no sex stereotyping, because the same question could have been asked of a father. Doing so implicitly chose an inference that the interviewer was thinking of parents generally rather than thinking of women with children. Judge Stahl noted but did not explicitly draw inferences from this question.

Chadwick’s most significant circumstantial evidence consisted of statements she alleged were made by Miller in explaining why Chadwick was not chosen for the promotion. “It was nothing you did or didn’t do. It was just that you’re going to school, you have the kids and you just have a lot on your plate right now.” In the same conversation, Miller told Chadwick that “if [the

---

57 550 F. Supp. 2d at 145.
58 Norman Stahl, Ballotpedia, https://ballotpedia.org/Norman_Stahl (last visited Sept. 25, 2018). Judge Stahl was appointed to the First Circuit to replace Justice David Souter when the latter was appointed to the U.S. Supreme Court.
59 550 F. Supp. 2d at 47 & n.10.
60 Id.
61 “The young age and unusually high number of children would have been more likely to draw the decisionmaker’s attention and strengthen any sex-based concern she had that a woman with young children would be a poor worker.” 561 F.3d at 47.
62 550 F. Supp. 2d at 143.
63 Id. at 143.
64 561 F.3d at 42.
three interviewers] were in your position, they would feel overwhelmed.”

Wellpoint took the position that the real reason Chadwick was denied the promotion was that she interviewed badly, and that Miller was just trying to offer a less hurtful reason for rejection. Judge Hornby recognized that the remark reflected discrimination against caregivers, but again would not recognize as reasonable an inference that it was based on sex:

If the case went to a jury on this record, the jury would have to speculate in order to reach a conclusion that Miller stereotyped working mothers and that she treated working mothers of young children worse than (given the opportunity) she would treat working fathers of young children. . . . Might Miller harbor such stereotypes? Yes: the jury might well suspect it. But suspicion is not enough; despite what might be the popular intuition about what Miller meant, I conclude that her use of sexual stereotypes cannot be shown by a preponderance of the evidence on this record.

For this judge, the missing link was something more specific proving that a man with four young children would not have been treated the same way, and he characterized Chadwick’s evidence as merely prevalent stereotypes plus a “sexually ambiguous utterance.” The plaintiff’s superior qualifications, the series of remarks, and the background of societal stereotyping was not enough to make an inference sufficiently believable compared to other possibilities.

Compare that to Judge Stahl’s opinion for the First Circuit, looking at the same record and applying the same standard of proof. It paid far more attention to “it was nothing you did or didn’t do,” the reference to “kids,” and the suggestion that the interviewers imagined Chadwick would be

---

65 Id.
66 550 F. Supp. 2d at 143. Miller so testified at her deposition, but in the summary judgment context both courts recognized that a jury, whose job it is to assess credibility, could choose to disbelieve her testimony. As the court of appeals noted, “Miller explained the non-promotion in one way to Chadwick (that she had too much on her plate with her kids and school) and in a very different way in her deposition (that Chadwick had performed poorly in her interviews). A jury could reasonably question the veracity of this second explanation given that Chadwick was an in-house, long-time employee who had worked closely with her interviewers, had received stellar performance reviews, and was already performing some of the key tasks of the Team Lead position. A jury could rightly question whether brief interviews would actually trump Chadwick’s apparently weighty qualifications, or whether, given the other circumstantial evidence discussed above, Chadwick was really passed over because of sex-based stereotypes.” 561 F.3d at 47 - 48.
67 Id. at 147 (emphasis added).
68 Id. at 147 & n.15.
“overwhelmed.” Putting this conversation together with the other circumstantial evidence (as read through the appellate court’s inferences), this opinion required no explicit comparison to males to infer that gender stereotypes, not just parenting stereotypes, were at work:

Given the common stereotype about the job performance of women with children and given the surrounding circumstantial evidence presented by Chadwick, we believe that a reasonable jury could find that Wellpoint would not have denied a promotion to a similarly qualified man because he had “too much on his plate” and would be “overwhelmed” by the new job, given “the kids” and his schooling.69

Once again, then, the courts’ opinions show an inference gap, and one that cannot be explained by the record itself. Two judges (the district judge and magistrate judge) found an inference of sex-based discrimination so unlikely that summary judgment was granted. Three judges (the First Circuit panel), reviewing the case de novo, found that such an inference would be proper, so that the case should go to the jury.70

B. Other Pretrial Rulings

Inferences that are contested but that end litigation are especially troubling. But the same phenomenon of judges’ reliance on arguable inferences also impacts non-dispositive pretrial rulings. Many such decisions provide wide discretion for the trial judge, and they are governed by general standards or multi-factor tests rather than strict rules. The decisions, though, will be based on facts as the judge finds them, whether explicitly or implicitly, consciously or subconsciously. Those forces create a context in which the judge’s beliefs about “facts” can play a powerful role when the governing standards are applied to them.

In making these rulings the judge is finding the facts and not merely deciding whether there is a genuine factual dispute for later jury

69 Id. at 48. The opinion also noted the “parallel stereotypes presuming a lack of domestic responsibilities for men.” The First Circuit affirmed the trial court’s decision to exclude a proposed plaintiff-side expert witness on sex stereotyping, but it did so on the narrower ground of the expert’s lack of knowledge about the facts of this particular case, a deficiency that presumably could have been remedied on remand.

70 On remand, the parties apparently settled the case in a way that included a judgment for defendant on stipulated facts that resembled defendants’ factual claims and that stipulates to a disagreement about what Miller said to Chadwick. Chadwick still worked for Wellpoint at the time of the stipulation. The relevant court documents are available through PACER and are on file with author.
determination. Except in very limited circumstances, the standard of proof for these pretrial rulings is the conventional “preponderance of the evidence” standard – more likely than not.\textsuperscript{71} The judge is undertaking a comparative task in evaluating information and reaching conclusions. And once again, it is the mental exercise of finding some inference more believable that is the source of differing outcomes. This section will highlight three examples: two are disputes about discovery,\textsuperscript{72} and one relates to class certification.\textsuperscript{73}

1. Relevance & Protective Orders: \textit{Cazorla v. Koch Foods}\textsuperscript{74}


\textsuperscript{72} Discovery relevance orders quite often turn on decisions about whether the information sought would support an inference in favor of the party seeking discovery. Balancing factors related to proportionality and requests for protective orders will also turn on the court’s belief about underlying facts to be inferred. Privilege decisions may turn on facts such as whether a document was created in anticipation of litigation, whether a lawyer was acting in her capacity as an attorney, or whether a communication was kept confidential, and many of those decisions will be inferences from circumstantial evidence. Even managerial decisions may turn on the judge’s beliefs about the strength of the inferential link between discovery and legal claims, the motivations of parties and attorneys, and the cost of discovery. For some recent examples of discovery decisions based on inferences, see \textit{generally} Wittmann v. Unum Life Ins. Co. of America, No. CV 17-9501, 2018 WL 3374164 (E.D. La. July 11, 2018) (inference from income sources to allegations of conflict of interest); Shah v. Metro. Life Ins. Co., No. 2:16-CV-1124, 2018 WL 2309595, at *10 (S.D. Ohio May 22, 2018) (inference from employee rewards and bonuses to their handling of plaintiff’s disability claim); English v. Washington Metro. Area Transit Auth., 323 F.R.D. 1 (D.D.C. 2017) (inferences from bus driver’s pre-accident activities to negligence in driving and dragging plaintiff under bus); Does I-XIX v. Boy Scouts of Am., No. 1:13-CV-00275-BLW, 2017 WL 3841902, at *3 (D. Idaho Sept. 1, 2017), reconsideration denied sub nom, Doe I-XIX v. Boy Scouts of Am., No. 1:13-CV-00275-BLW, 2017 WL 6029591 (D. Idaho Dec. 5, 2017) (inference between content of Boy Scout files and claims of misrepresentation and abuse); Bias v. Tangipahoa Par. Sch. Bd., No. CV 12-2202, 2017 WL 679365, at *5 (E.D. La. Feb. 21, 2017) (inference between plaintiff’s military records, transfer, and claim of retaliation for filing False Claims Act suit); Centeno v. City of Fresno, No. 116CV00653DADSAB, 2016 WL 7491634, at *5 (E.D. Cal. Dec. 29, 2016) (inference between prior complaints about defendant officers and actions in current excessive force claim).

\textsuperscript{73} The requirements for class certification turn on fact findings, many of which in turn are based on inferences. \textit{See}, e.g., Menocal v. GEO Grp., Inc., 882 F.3d 905, 918 (10th Cir. 2018) (finding the predominance requirement for a 23(b)(3) class action met when plaintiffs will be able to prove causation through class-wide inference); Arkansas Teachers Ret. Sys. v. Goldman Sachs Grp., Inc., 879 F.3d 474, 483 (2d Cir. 2018) (recognizing that “fraud on the market” theory is a presumption that an inference can be drawn); Waggoner v. Barclays PLC, 875 F.3d 79 (2d Cir. 2017), \textit{cert. denied}, 138 S. Ct. 1702 (2018) (discussing factors from which inference of efficient market may be drawn); Rikos v. Procter & Gamble Co., 799 F.3d 497, 512–13 (6th Cir. 2015) (finding that materiality of misrepresentations could support an inference of reliance for entire class).

\textsuperscript{74} Cazorla v. Koch Foods of Mississippi, LLC, No. 3:10cv135-DPJ-FKB, 2014 WL
Hispanic employees of Koch Foods, a poultry processor, alleged that they were subjected to harassment and abuse on the job. Koch, on the other hand, claimed that the employees had made up the allegations in order to get U visas, which are available to abuse victims who assist in government investigations.

Those opposing positions led to a discovery dispute: the company sought discovery of any information related to the employees’ U visa applications. The plaintiffs objected, arguing that the discovery would reveal to Koch the immigration status of any applicants and their families, creating risks of job loss and deportation. The magistrate judge and district judge allowed U-visa-related discovery but entered a protective order prohibiting certain uses of the information. The Fifth Circuit vacated the trial court’s order.

Koch Foods operates a large poultry plant in Morton, Mississippi, and the dispute underlying this lawsuit arose out of allegations about harassment in the room where approximately 85 employees debone and package chicken thighs. The workers in this room were almost all Hispanic, most were illiterate and spoke little English, and many were undocumented. Here are the workers’ allegations, as summarized by the Fifth Circuit:

Koch supervisors allegedly groped female workers, and in some cases assaulted them more violently; offered female workers money or promotions for sex; made sexist and racist comments; punched, elbowed, and otherwise physically abused workers of both sexes; and demanded money from them in exchange for permission for bathroom breaks, sick leave, and transfers to other positions. When workers complained or resisted, Koch managers allegedly ignored them, and some debone supervisors allegedly retaliated by docking their pay; demoting, reassigning, or firing them; and threatening to physically harm them or have them arrested or deported.

Ten workers filed the first claims with the EEOC in 2009, and they and a few more followed up with lawsuits in 2010 and 2011. Meanwhile in 2010, the EEOC launched its own investigation, found reasonable cause to believe a violation had occurred, and filed its own suit identifying a class of about 50-75. The cases were consolidated, and several additional employees
Koch denied the allegations, arguing that all of the complainants made up their charges to qualify for improved immigration status under the federal U-visa program. It is that denial defense that led to the discovery requests at issue here. The claimants requested a protective order under Rule 26(c), arguing that the information sought was not relevant, and that even if it was relevant, the benefit of the information to Koch was outweighed by the harm that compliance would cause to the claimants and others.

The magistrate judge, district judge, and Fifth Circuit panel all addressed the same balancing test in evaluating the defendant’s discovery request and plaintiffs’ request for a protective order: did the potential relevance of the U-visa information outweigh the in terrorem effect of revealing that immigration-relevant information to Koch? In making that decision, the trial level judges and Fifth Circuit judges differed in their factual beliefs about both relevance and the fears held by immigrant populations.

Implementation of the Rule 26(c) balancing test required subsidiary fact-finding in order to determine the extent of relevance and harm. At the trial level, with respect to relevance, the magistrate judge and district judge found the complainants’ credibility to be a central issue and the immigration information highly relevant to assessing it. Those judges concurred with Koch’s argued inference that the existence of an “exponential jump” or

77 Id.
78 The U-visa program offers four years of nonimmigrant status to qualifying individuals and their family members and allows them to apply for green cards. Those with pending U-visa applications may also attain work authorizations. The program is available to victims of substantial physical or mental abuse who aid a law enforcement agency in investigating the alleged offenses. That agency must certify that the applicant is aiding the investigation, and the U.S. Customs and Immigration Service must conduct its own de novo review of relevant evidence and confirm the victim’s eligibility. 838 F.3d at 545.
80 In considering many discovery disputes, the court is using the information at hand to predict what will happen in the future rather than assessing the comparative probability of past events. Nevertheless, the factors that underlie discovery orders are very fact-based. The Advisory Committee Notes to the 2015 amendments to Rule 26 provide that at the trial level, “[t]he court’s responsibility, using all the information provided by the parties, is to consider these and all the other factors in reaching a case-specific determination of the appropriate scope of discovery.” 2014 WL 12639863 at *5. The applicable standard of review was deferential even at the trial level. The magistrate judge’s opinion could only be changed if it was “clearly erroneous” or “contrary to law.” FED. R. CIV. P. 72(a). With minor exceptions, the district judge approved of the magistrate judge’s order.
“spike” in claims after the EEOC became involved, coupled with applications for U visas, was relevant to the issue of the claimants’ credibility and thus to the truth of the underlying claims.

With respect to prejudice if discovery were allowed, the trial judges refused to infer from the claimants’ immigration and employment situation that being forced to reveal the U-visa applications to Koch would have an in terrorem effect on the claimants themselves or on other who might find themselves in a similar situation. First, the trial judges found, none of the named plaintiffs work for Koch anymore, so they would not fear being fired. Further, inferred the trial judges, claimants who had applied for U visas had already revealed to federal immigration authorities that they were not properly in the United States, so even if Koch reported them to immigration officials there could be no additional harm. At the trial level, then, the judges discounted the claimants’ statements of their own subjective fears and did not infer future harm from disclosure to Koch.

The Fifth Circuit judges, in an opinion written by Judge Patrick Higginbotham, might well not have vacated the order based only on private interests, given the very deferential “abuse of discretion” standard of review. They did agree that the information sought was relevant for purposes of impeachment. But their inferences were different. For example, the Fifth Circuit found that the “spike in claims” number, on its own, was “not particularly suggestive of mass fraud.” Instead, for example, “the EEOC may have discovered additional harassment claimants during the pre-suit conciliation and investigation processes.” The appellate judges relied on information about statutory provisions that deter false U visa claims, and also cited evidence suggesting a pattern of abuse of immigrant workers in the

---

82 According to the district court, only eight people initially filed claims with the EEOC while in its Third Amended Complaint the EEOC identified 115 claimants, 44 of whom were women who claimed to have been sexually harassed and the rest of whom claimed some other kind of harassment or injury. 2014 WL 12639863 at *5.
84 As that opinion notes, the trial court will be found to have abused its discretion in balancing only “when a relevant factor that should have been given significant weight is not considered; when an irrelevant or improper factor is considered and given significant weight; and when all proper factors, and no improper ones, are considered, but the court, in weighing those factors, commits a clear error of judgment.” 838 F.3d 547 (quoting Kern v. TXO Prod. Corp., 738 F.2d 968, 970 (8th Cir. 1984)). A finding of abuse of discretion could thus be based on the evidence considered rather than the inference itself. Nevertheless, as discussed above, the Fifth Circuit opinion in this case reveals that its inferences differ from those of the trial court.
85 838 F.3d at 558.
poultry industry.86

Other differences relate to the “prejudice” side of the balancing test. The appellate judges inferred that disclosure of the U-visa information would cause significant harm, both to the individual claimants and to their family members. With respect to employment, the Fifth Circuit judges believed that the evidence showed “a risk that U visa discovery will cause some claimants or family members to lose their jobs.”87 Nor did the appellate judges conclude that disclosure to U-visa processors removed fear of all immigration consequences. Instead, it found that “[a]n abuse victim might well be willing to disclose sensitive information to a few sympathetic officials, yet nonetheless fear that his or her abuser might obtain that information and spread it far and wide. . . . [T]heir having submitted U visa applications does not rule out an in terrorem effect from further disclosure.”88

The factual disagreement about the impact of disclosure once again comes from a difference in which facts the different judges chose to consider as well as the inferences to be drawn from them. In addition to discovery’s potential impact on the complainants, Judge Higginbotham’s opinion considered the harm that disclosure would cause to the enforcement efforts of the EEOC, amicus NLRB, and law enforcement agencies nationwide, finding that disclosure would deter immigrant victims of abuse, thereby frustrating Congress’s purpose in creating the U-visa program.89

Putting all of those factual conclusions together (lesser relevance, stronger individual harm, and very strong societal harm that had not been considered by the trial court), the Fifth Circuit vacated the trial level discovery order and remanded for further proceedings.

2. Relevance & Burden: Kuttner v. Zaruba90

This discovery dispute arose out of an employment discrimination case. The difference in inferences here are found in disagreements on appeal between the majority opinion of Judges Sykes and Easterbrook and the dissent written by Judge Posner.

---

86 Id.
87 Id. at 561.
88 Id.
89 Id. at 563-64.
90 Kuttner v. Zaruba, 819 F.3d 970 (7th Cir. 2016) and 819 F.3d at 977 (Posner, J., dissenting).
Plaintiff Susan Kuttner worked for the DuPage County Sheriff’s office beginning in 1998. In 2010, she was fired based on a claim that she had worn (at least part of) her uniform while trying to collect a loan for her boyfriend. She visited the debtor’s home, spoke to his father (the debtor himself was not at home), and left her business card. Kuttner admitted to violating two rules: “conduct unbecoming” an officer and improper wearing of her uniform, and as a result Sheriff Zaruba fired her. Kuttner filed a sex discrimination suit, alleging that male deputies had received less severe punishment for more severe violations on a number of occasions.91

Kuttner sought to discover the personnel files of 30 named deputies who she believed had violated department policies. The trial judge limited that discovery to incidents after January 1, 2006 and to charges involving abuse of authority. Toward the end of the discovery period, the judge also refused to allow Kuttner’s lawyer to ask questions on deposition about what deponents had heard about violations, limiting the inquiry to the deponent’s personal knowledge of infractions. Operating under those limitations, Kuttner was unable to convince the trial judge that un-fired male deputies’ conduct was comparable, and so granted summary judgment to Sheriff Zaruba.92

On appeal, a split Seventh Circuit panel affirmed. Judge Sykes,93 writing for the majority, held that the district judge had not abused his discretion in denying discovery.94 The majority found that the passage of time made inferences that comparable male deputies had not been as severely disciplined too attenuated to be relevant. The decision not to include the first seven years of Kuttner’s employment “served to hone in on possible comparators who were reasonably likely to have been subject to the same rules, supervisors, and decisionmaking process as Kuttner.”95 Having affirmed the denial of discovery, the majority also affirmed the grant of summary judgment. Finding only infractions involving “improper projection of coercive police authority in service of a personal end” to be sufficiently

---

91 819 F.3d at 972.
92 Id. at 973.
93 Diane Sykes, Ballotpedia, https://ballotpedia.org/Diane_Sykes (last visited Dec. 10, 2018). Judge Sykes was on President Donald Trump’s list of 25 potential Supreme Court nominees to replace Justice Anthony Kennedy on the Court. Id.
94 Id. at 974 (“This standard requires us to affirm unless the judge’s ruling lacks a basis in law or fact or clearly appears to be arbitrary.”)
95 Id. at 974. Judge Posner’s dissent, however, pointed out that “the sheriff’s lawyer acknowledged that nothing had changed in 2006, and so there was no reason to ignore the earlier misconduct, which was the critical evidence of discriminatory treatment of the plaintiff.” Id. at 979.
similar to justify an inference of discrimination, the majority rejected the four post-2006 examples Kuttner brought to light.96

Judge Posner,97 on the other hand, would have found the denied discovery about other employees to be not merely relevant but also so significant that the trial court’s action required reversal. Kuttner listed twenty-one examples in her complaint, and Judge Posner highlighted ten of them, including sexual misdeeds, abuse of power, and domestic violence, some committed while in uniform.98

In Judge Posner’s view, all of these officers’ misconduct (none of which resulted in firing) were at least as serious as the plaintiff’s actions and, because of inferences of discrimination that could result from the comparison, relevant for purposes of discovery. He wrote, “It is a virtual certainty that the plaintiff was disciplined far more harshly than male counterparts who engaged in far more egregious conduct – far more harshly because she’s a woman. The DuPage County Sheriff’s Office is or at least was a boy’s club.”99

Note, then, the difference in decisions about what the information sought might be able to show. Judge Sykes’s majority opinion wanted to see specific parallels in time and circumstances before the trial court could draw an inference of discrimination-based disparate treatment. Under this view of the potential probative value (relevance) of other incidents, the denial of discovery was supported by the facts. Judge Posner, on the other hand, thought the trial court should have found the information relevant because a wider array of disciplinary decisions could form the basis for an inference of discrimination.

3. Class Certification: Mielo v. Steak ‘N Shake100

This putative class action against Steak ‘N Shake alleged that the company violated Title III of the Americans with Disabilities Act (“ADA”) because their parking lots are not fully accessible to and independently usable by individuals who use wheelchairs or are otherwise mobility disabled. This

96 Id. at 976-77.
98 Id. at 978. The list contains numerous salacious and disturbing details.
99 Id.
time, the trial court reached a plaintiff-friendly result, and the court of appeals reversed. In the trial court, Magistrate Judge Robert C. Mitchell\textsuperscript{101} certified a broadly defined class.

The Third Circuit, in an opinion written by Chief Judge D. Brooks Smith,\textsuperscript{102} based its reversal in part on a disagreement about the class certification requirement of numerosity, which in turn was based on the judges’ different views of inferences.\textsuperscript{103} While there is no exact number that satisfies numerosity, courts generally state that approximately 40 class members are enough.\textsuperscript{104} Under Third Circuit precedent, class certification “calls for a rigorous analysis in which ‘[f]actual determinations supporting Rule 23 findings must be made by a preponderance of the evidence.’”\textsuperscript{105} So what was the evidence that did or did not support an inference that the class was so numerous that joinder would be impracticable?\textsuperscript{106}

The trial court was persuaded that the numerosity requirement was met by these pieces of information: 1) census data showing that there are between 14.9 million to 20.9 million persons with mobility disabilities who live in the United States; 2) the statement of a Steak ‘N Shake executive that “thousands of people with disabilities utilize [Steak ‘n Shake] parking lots . . . each year”; 3) plaintiffs’ identification of eight locations in Pennsylvania and Ohio that had noncompliant parking lots; 4) the existence of more than 400 Steak ‘N Shake restaurants in the United States potentially subject to the class’s claims; and 5) the special difficulties that arise because the group of putative class members includes a potentially high number of individuals with mobility disabilities from multiple states. The plaintiffs urged the court to use its “common sense” to conclude that the numerosity requirement was

\begin{footnotes}
\footnote{Brooks Smith, Ballotpedia, https://ballotpedia.org/Brooks_Smith (last visited Sept. 16, 2018).}
\footnote{897 F.3d at 487. The court of appeals judges also concluded, based on the many ways in which the restaurant facilities might violate Title III of the ADA, that the “commonality” requirement of Rule 23(a) was not met. 897 F.3d at 487-490. The district court had found the requirement to be met based on the inferences from facts in the record: that Steak ‘N Shake applied the same policies on parking lot access nationwide. 2017 WL 1519544 at *6. Their disagreement, however, turned less on inferences and more on breadth of the district court’s definition of the class, which could be read to include all types of “accessibility barriers” despite the definition’s reference to parking facilities.}
\footnote{Stewart v. Abraham, 275 F.3d 220, 226–27 (3d Cir. 2001).}
\footnote{897 F.3d 483-84, quoting In re Hydrogen Peroxide Antitrust Litigation, 552 F.3d 305, 307 (3d Cir. 2009).}
\footnote{Fed. R. Civ. P. 23(a)(1).}
\end{footnotes}
satisfied, and it did so.107

The Third Circuit judges reached the opposite conclusion. Mentioning the census data and characterizing the executive statement as a “single off-hand comment . . . speculating [about numbers],” those judges believed that what the plaintiffs saw as common sense was really “mere speculation.” Their opinion states:

Plaintiffs’ first strand of evidence—indicating that there are between 14.9 million to 20.9 million persons with mobility disabilities who live in the United States—suggests that it is highly likely that at least 40 of those individuals would have experienced access violations at one of the Steak ‘n Shake locations at issue in this litigation. But although those odds might be enough for a good wager, we must be mindful that “[m]ere speculation as to the number of class members—even if such speculation is ‘a bet worth making’—cannot support a finding of numerosity.”108

Without more direct or specific evidence of the number of disabled individuals who actually patronized a Steak ‘N Shake restaurant and experienced an ADA violation there, the judges on the Third Circuit found an inference of numerosity to be unsupported by the evidence.109 Although it employed an abuse of discretion standard of review,110 the court had noted that plaintiffs seeking class certification must prove the facts supporting the Rule 23(a) requirements by a preponderance of the evidence.111 The existence of at least forty class members was not, to those judges, more believable than a contrary inference.112

V. JUDGES AND HEURISTICS

107 Id. at 485-86; 2017 WL 1519544 at *5-6.
108 897 F.3d at 486 (emphasis added) (quoting Hayes v. Wal-Mart Stores, Inc., 725 F.3d 349, 357 (3d Cir. 2013)).
109 The appellate panel’s decision may also have been influenced by its dim view of the merits. “Despite the novelty of these interpretations, Steak ‘N Shake has not yet filed a motion to dismiss or motion for summary judgment.” 897 F.3d at 475.
110 Id. at 474.
111 Id. at 484.
112 “If Plaintiffs wish to attempt to satisfy their Rule 23(a)(1) burden upon remand, they will need to provide evidence that will permit the District Court to conclude that a sufficiently numerous group of disabled individuals have experienced or will experience ADA violations at a relevant Steak ‘n Shake restaurant, and that joinder is thereby impracticable.” Id. at 487. The Third Circuit also reviewed de novo the district judge’s legal framework, and rejected his “when in doubt, certify the class” approach. Id. at 483-84.
The hypothetical in section III and real rulings in Section IV show judges drawing different inferences based on the same underlying facts. This section will describe the empirical literature showing the ways in which mental shortcuts (“heuristics”), which can vary with individuals’ experiences and attitudes, may explain why inferences can vary so dramatically from judge to judge. Not surprisingly, the literature does not test all heuristics or all types of decisions, but this sample shows that judges’ expertise does not usually protect them from the operation of intuitive thought processes.

A. Introduction to Heuristics: Unconscious Shortcuts

Judges’ generalizations operate through the reasoning mechanisms that human brains use. Like most cognition, judicial fact finding often happens under circumstances of incomplete information and limited processing time.113 Psychologists use a number of different labels to describe how humans think, but they recognize distinct roles for intuition and deliberation.114 Intuitive processes (sometimes called System 1) operate quickly, often subconsciously, are relatively effortless, and are susceptible to emotional influences. “System 1 is radically insensitive to both the quality and quantity of the information that give rise to impressions and intuitions.”115 Deliberative processes (System 2), on the other hand, move more slowly because they require conscious work and the application of rules. The two systems interact, and the intuitive inferences, which are automatically gathered, affect more deliberative decisions in a number of ways. Despite the intervention of System 2 thinking, the influence of System 1 rarely ceases.116

System 1 employs various heuristics to help make those quick evaluations. Although accurate and helpful in many situations, they can also lead to unreliable conclusions.117 Keep in mind that System 1 heuristics

113 James R. Steiner-Dillon, Epistemic Exceptionalism, 52 Indiana L. Rev. – (MS p. 18) (forthcoming 2019), available at https://papers.ssrn.com/abstract=3130745 (last revised Aug. 5, 2018) (noting that the “empirical literature on judicial susceptibility to cognitive biases shows that judges’ training and experience leave them better prepared to resist some forms of cognitive error, but on the whole, does not diverge from the cognitive processes of laypersons”).
114 Guthrie, Blinking, at 6 (discussing the judicial decision-making process). As Guthrie notes, “The convergence of psychologists on the notion that two separate systems of reasoning coexist in the human brain is remarkable.” Id. at 9 n.49. See also Daniel Kahneman, Thinking, Fast and Slow 4 (2011) (discussing the mental process of impression, intuition, and decision).
115 Kahneman, supra note xx at 86.
116 Kahneman, supra note xx at 85-86.
117 See, e.g., Amos Tversky & Daniel Kahneman, Judgment Under Uncertainty:
operate automatically – these mental processes are intuitive and powerful, affecting thought in a way that is hidden from the conscious, deliberate System 2. Its efficient operation also means that it does not keep track of discarded possibilities – System 1 loves coherence and makes available a collection of information that fits together and is compatible with the person’s overall worldview.

This sub-section will highlight some of the heuristics that are relevant to judicial inferences and whose impact on judges has been studied.

**Availability.** We make judgments about how likely something is based on how easily examples come to mind—how “available” they are.\(^\text{118}\) That could be based on how many examples one has encountered, and to that extent it is related to frequency which is related to probability. But it is also affected by other things that make something “available” (memorable): things like dramatic content, personal impact, and visual stimulation. The availability heuristic can make people think sharks kill more people than ponies, or that they are more likely to die from a terrorist attack than from heart disease. The operation of the availability heuristic for any particular individual will be influenced by that person’s experiences, since that impacts what information is “available” to System 1’s quick search.

**Representativeness.** The “representativeness” heuristic causes people to estimate probability using the similarity of one thing to the person’s mental image of the corresponding stereotype.\(^\text{119}\) When people rely on representativeness to make judgments, they are likely to judge wrongly because the fact that something is more representative does not actually make it more likely. The representativeness heuristic can lead “to serious errors, because similarity, or representativeness, is not influenced by several factors that should affect judgments of probability.”\(^\text{120}\) For example, it is insensitive to the accuracy (or lack thereof) of the stereotype. And stereotypes are influenced by the person’s own experiences and understanding of the way the world works.

**Confirmation.** This heuristic is defined as “the seeking or interpreting of evidence in ways that are partial to existing beliefs, expectations, or a

---

\(^{118}\) *Heuristics and Biases*, 185 SCIENCE 1124 (1974) (discussing biases resulting from use of representativeness heuristic, availability heuristic, and anchoring heuristic).

\(^{119}\) *Kahneman*, supra note xx at x.

\(^{120}\) *Id.* at x. See also Tversky & Kahneman, *Uncertainty*, supra note xx at 1124-25.
hypothesis in hand.” It operates as part of System 1 thinking. Confirmation bias leads people to find and interpret information in a way that supports preexisting hypotheses and to avoid information or interpretations that support alternate possibilities. It can also take the form of giving greater weight to information supporting a position one has taken or remembering that supporting information more readily than information that disconfirms the belief. This is particularly common in situations that are inherently complex and ambiguous—those that are characterized by interactions among numerous variables and in which the cause and effect relationships are obscure. Like the availability heuristic, the operation of any particular person’s confirmation bias very much turns on that person’s experience and view of “common sense.”

Affect. The affect heuristic leads people to decisions that are consistent with their emotions. “Do I like it? Do I hate it? How strongly do I feel about it?”

Emotions influence how people perceive others, what they remember, about others, and how they process information about others. Emotions guide “people’s attitudes, beliefs, and inferential strategies’ so that they see people they like as having positive qualities and people they do not like as possessing negative ones. Consequently, even deliberative reasoning can be influenced by intuitive, emotional reactions.”

System 2 is particularly ineffective in overcoming the affect heuristic.

Anchoring. When people estimate numerical values, they often rely on an initial value available to them and adjust from there—hence the anchor metaphor. This can be rational and helpful. But it turns out that the heuristic is at play even when it makes no sense at all, such as when the number is unrelated to the number to be estimated.

121 Raymond S. Nickerson, Confirmation Bias: A Ubiquitous Phenomenon in Many Guises, 2 REV. GEN’L PSYCHOL. 175, 175 (1998).
122 Id. at 191-92.
123 KAHNEMAN, supra note xx at 139.
125 KAHNEMAN, supra note xx at 103 - 04.
127 Chris Guthrie et al., Inside the Judicial Mind, 86 CORNELL L. REV. 777, 788 (2001)
Another form of anchoring is not about numbers. Everyday social interactions (as well as judicial decisions) require humans to infer the thoughts and feelings of other people. In order to figure this out, people use themselves as a kind of “anchor” to estimate the intent and emotions of others and adjust from there depending on how similar or dissimilar from themselves the other person seems to be. “For example, to infer another’s mental state, one might first mentally imagine experiencing that person’s situation, read off the evoked mental state, and then assume that the other person would feel similarly.”\textsuperscript{128} This process, sometimes called “self-referencing,” works very much like anchoring with numbers.

When people use themselves as an anchor point, if the other person’s experiences are different, the inference will be unreliable unless it is adjusted to account for dissimilarities. (The adjustment may use System 2 deliberative thinking.) This can work relatively well so long as the “other” is fairly similar and sufficient adjustments are made. However, one study found that when the “other” is not perceived as similar, the people doing the inferring do not use self-referencing as an anchor. Instead, they fall back on stereotypes.\textsuperscript{129}

\textit{Hindsight.} “[O]nce people know the outcome of an event, they tend to overestimate what could have been anticipated in foresight.”\textsuperscript{130} Experimenters have detected hindsight bias in a wide array of situations, including “general-knowledge questions, in political or business developments, in predictions of elections or sport results, in medical diagnoses or in personality assessment,” and more.\textsuperscript{131}

\textit{Egocentric heuristics and biases.} In addition to self-referencing, a cluster of self-centered biases can also affect judicial decisions. One is

\textsuperscript{128} Diana Tamir & Jason Mitchell, \textit{Anchoring and Adjustment During Social Inferences}, 142 \textit{J. EXPERIMENTAL PSYCH: GEN.} 151 (Apr. 16, 2012), available at \url{http://dx.doi.org/10.1037/a0028232}.


\textsuperscript{130} \textit{Hindsight Bias}, \url{https://psychology.iresearchnet.com/social-psychology/decision-making/hindsight-bias} (last visited Oct. 29, 2018).

\textsuperscript{131} Id.
“egocentric bias,” a tendency to have a higher opinion of oneself than reality would support. This can make people overconfident in the accuracy of their opinions.\(^{132}\) A second is the “false consensus effect,” in which people overestimate the degree to which others agree with their beliefs, which also creates overconfidence.\(^{133}\) Those biases are related to “naïve realism,” the human tendency to believe that we see the world around us objectively, and that people who disagree with us must be uninformed, irrational, or biased.\(^{134}\) One consequence of naïve realism is referred to as the “bias blind spot,” which is the ability to recognize cognitive and motivational biases in others while failing to recognize the impact of bias on the self.\(^{135}\) Sadly, when people form beliefs based on heuristics they may well be both confident that the decisions are correct and unaware of the impact of those heuristics on the decisions.

\section*{B. Heuristics Affect Decisions}

It would be tempting to believe that judges, because of their high intelligence, extensive education, and professional experience would not be subject to heuristic-based reasoning flaws. Sadly, this is not true. Some of the evidence for that is circumstantial. Research on other types of professional experts – including doctors, real estate agents, psychologists, options traders, military leaders, and auditors – show that they rely on heuristic biases.\(^{136}\) Even more tellingly, empirical studies of actual judges show that heuristics influence judicial tasks.

Many of the contemporary studies of trial court decision-making have been done by the trio of Professors Jeff Rachlinski and Chris Guthrie and U.S. Magistrate Judge Andrew Wistrich (hereinafter RGW). Overall, those studies of hundreds of federal and state judges show that “judges rely on heuristics that can lead to systematically erroneous judgments.”\(^{137}\)

\begin{footnotesize}
\begin{itemize}
\item[^{135}] Emily Pronin et al., \textit{The Bias Blind Spot: Perceptions of Bias in Self Versus Others}, 28 PERSONALITY & SOC. PSYCHOL. BULL. 369–381 (2002).
\item[^{136}] Jeffrey J. Rachlinski et al., \textit{Inside the Bankruptcy Judge’s Mind}, 86 BOSTON U. L. REV. 1227, 1229-30 (2006); Guthrie, \textit{Inside the Judicial Mind}, \textit{supra} note xx at 783.
\item[^{137}] Guthrie, \textit{Judicial Mind}, \textit{supra} note xx at 784.
\end{itemize}
\end{footnotesize}
Administering a “cognitive reflection test” to judges showed that “judges tended to favor intuitive rather than deliberative faculties.”138 That test, however, asks questions unrelated to judging.139 Judicial task case studies also showed that judges tend to be influenced by heuristics such as anchoring, hindsight bias, and egocentric bias.140 Further, tests showed that judges have difficulty deliberately disregarding inadmissible evidence.141

To make this more concrete, consider some examples. First, in order to test whether judges are affected by anchoring, one RGW study presented participating judges with a hypothetical lawsuit arising out of an automobile accident. The control group judges were given the facts and asked, “How much would you award the plaintiff in compensatory damages?” Judges in the anchor group were given identical information, but they were also told of a motion to dismiss for lack of subject matter jurisdiction arguing that the amount in controversy was lower than the required $75,000. Judges were asked to rule on the motion and then, if they did not dismiss the case, asked the same question about compensatory damages.142 The judges’ responses showed that the $75,000 served as an anchor: judges who had not ruled on the motion (no anchor) awarded the plaintiff an average of $1,249,000, but judges who had been exposed to the motion awarded the plaintiff an average of only $882,000. The difference was statistically significant; the average award of the anchor group was 30% lower. “The anchor triggered intuitive, automatic processing that the judges were unable to override.”143 The judges’ inferences about damages were influenced by the anchor.144

A test of hindsight bias informed judges about the outcome on appeal of a case involving Rule 11 sanctions. The judges were all given the same fact pattern, except that they were given one of three different appellate outcomes. They were then asked, in light of the fact pattern and all three possible outcomes, which of the outcomes had been the most likely. In each

---

138 Guthrie, Blinking, supra note xx at 17.
139 Id. at 10 (describing CRT).
140 Id. at 19-27.
142 Because the plaintiff as described clearly had incurred damages exceeding $75,000, the researchers regarded the motion to dismiss as meritless, and only 2 out of 116 judges would have granted it. Id. at 21.
143 Id.
144 In another hypothetical, judges were either told of a $10 million settlement demand or just told that the plaintiff “was intent on collecting a significant monetary payment.” The $10 million acted as a high anchor, resulting in a significant increase in the judges’ damage awards despite otherwise identical fact patterns. Id. at 1288-92.
case judges found the outcome they had originally been given to be significantly more likely. “Learning an outcome clearly influenced the judges’ ex post assessments of the ex ante likelihood of various possible outcomes. The intuitive notion that the past was predictable prevailed.”

In an RGW study designed to test judges’ ability to disregard inadmissible evidence, both control group and “suppression” judges were given information about available evidence in a contract dispute between Jones, a freelance consultant, and SmithFilms, a movie studio. The hypothetical lawsuit turned on whether the studio had offered Jones “producer credit.” Jones claimed that it had; the studio denied it. The only writing in the case was a short letter specifying that Jones would receive “other consideration as agreed upon by the parties during the pre-signing breakfast.” To corroborate his own testimony, Jones offered the testimony of a waitress who said she overheard a mention of producer credit. Smith was in a coma (!), so SmithFilms could offer only testimony that SmithFilms did not generally offer producer credit.

Both groups of judges were given this description and asked whether they would rule for Jones or SmithFilms. But the judges assigned to the “suppression” group first had to rule on the admissibility of an audiotape of a post-negotiation conversation between Jones and his attorney, with Jones objecting that the tape was protected by the attorney-client privilege (and under the facts of the hypothetical, that was the more likely conclusion). On the tape, Jones admits that he had not actually asked for the producer credit and there was no such agreement. Although the judges otherwise had identical information, the judges’ awareness of the plaintiff’s admission significantly decreased his win rate. “In the control condition [no tape], 55.6% (25 out of 45) of the judges found for the plaintiff. . . . Among those who ruled that the audiotape was privileged, only 29.2% (7 out of 24) found for the plaintiff.” Even judges who had ruled the tape out of bounds, then, let the knowledge they gained from the tape influence the inferences on which they based their judgments.

The RGW researchers also examined whether judges were subject to the affect heuristic, using a number of hypotheticals on hundreds of judges over a period of years. While judges no doubt strive to avoid the influence of

---

145 Guthrie, Blinking, supra note xx at 26. However, in another test of hindsight bias the judges were not affected by learning about the fruits of an arguably illegal search. See text accompanying notes xx infra.
146 Wistrich, Can Judges, supra note xx at 1294 – 95.
147 Id. at 1296.
emotions or non-merits values on their decisions, the results again show the impact of intuitive processes. For example, the researchers gave a problem involving illegal immigration to 508 judges. The judges were told that they were presiding over the prosecution of an immigrant illegally in the country. The defendant was a Peruvian citizen who pasted a forged U.S. entry visa into his legitimate Peruvian passport. The judges were told that the defendant was charged with “forging an identification card,” and given the language of the relevant statute. The statute’s application to these facts was open to interpretation.

The judges were divided into two groups. One was told that the defendant was a killer who had been hired by a drug cartel to sneak into the U.S. to track down someone who had stolen drug proceeds. The other group was told that the defendant was a father who had tried to sneak into the U.S. to earn money so he could pay for a liver transplant to save the life of his critically ill nine-year-old daughter. There was, therefore, a huge difference in the sympathy one would likely feel for the defendant—but not a difference relevant to whether pasting a forged visa into a legitimate passport constituted forging an identification card.

Did it affect the judges’ decisions? Yes, and pretty dramatically. “Among the judges who reviewed the father version, [only] 44% (102 out of 234) ruled that the act constituted forgery, as compared to 60% (154 out of 257) out of the judges who reviewed the killer version.”

C. When Does Professional Expertise Intervene?

Although they have not tested all possible heuristics, it seems clear from the RGW studies that judges as a group unwittingly employ intuitive thinking on a regular basis. But their research also shows that occasionally, professional training and habits of mind allow judges to override the misleading influence of System 1. This could happen in a couple of ways.

First, under some circumstances experts can develop such extensive, specific knowledge and skills that a mental operation that would normally be

148 Andrew J. Wistrich et al., Heart Versus Head: Do Judges Follow the Law or Follow Their Feelings, 93 TEX. L. REV. 855, 877 (2015).
149 Id. at 877-88.
150 Id. For another study showing a powerful unacknowledged influence of defendants’ characteristics on judicial reasoning, see Holger Spamann & Lars Klohn, Justice is Less Blind, and Less Legalistic than We Thought: Evidence from an Experiment with Real Judges, 45 J. LEGAL STUDIES 255 (2016).
a System 2 (deliberative) task becomes automatic and intuitive.\textsuperscript{151} Chess grandmasters, for example, develop such a deep sense of the game that they intuitively recognize patterns.\textsuperscript{152} However, it is unlikely that judges often acquire a level of internalized reflection that would protect them from the operation of heuristics. The conditions under which judges work (complex, varying decisions and limited feedback)\textsuperscript{153} are a serious impediment to developing the kind of deep but automatic pattern recognition that helps chess masters.

Second, judges’ training and practice in legal reasoning seems to allow System 2 deliberative thinking to overcome the influence of heuristics when the decision in question is about law application and is governed by a web of quite specific rules (as compared to discretionary decisions, general standards, or balancing tests). In one RGW test, designed to see if judges tend to err based on the hindsight heuristic, all of the judges were given the same core set of facts:

A police officer was on patrol outside a rock concert. The officer saw a well-dressed, nervous-looking man exit a BMW and fiddle with something in the trunk before he entered the concert. A half hour later, the officer noticed that one of the BMW’s windows was down. Concerned that someone might burglarize the car, he approached to close the window. Upon reaching the car, he smelled something that he believed, based on a demonstration at a training session several years earlier, to be burnt methamphetamine. He looked inside the car and didn’t see any drugs, but he did notice some Visine, a local map, and a couple of empty beer cans.\textsuperscript{154}

Participating judges were divided into two groups. The “foresight” judges were asked to say, based on these facts, whether this information supplied sufficient probable cause that they would issue a telephonic search warrant when called by the officer. The “hindsight” judges, however, were told that the police officer “conducted a warrantless search of the trunk and

\textsuperscript{151} ROBIN HOGARTH, EDUCATING INTUITION 204 (2001) (“Many processes or reactions to stimuli that once relied heavily on the deliberate system can over time become automatic and thus bypass consciousness. This migration from the deliberate system to the tacit [system] is an important characteristic of the phenomenon of expertise.”)

\textsuperscript{152} Daniel Kahneman & Shane Frederick, Representativeness Revisited: Attribute Substitution in Intuitive Judgment, in HEURISTICS AND BIASES: THE PSYCHOLOGY OF INTUITIVE JUDGMENT, at 51 (2002) (“A demonstration of the intelligence of System I is the ability of chess masters to perceive the strength or weakness of chess positions instantly.”)

\textsuperscript{153} Guthrie, Blinking, supra note xx at 32.

\textsuperscript{154} Id. at 26-27
found ten pounds of methamphetamine, other drug paraphernalia, and a recently fired gun that had been used earlier in the day to murder a drug dealer across town.” Judges in that group were asked to rule on a motion to suppress the evidence found in that warrantless search because the officer lacked probable cause.

If the hindsight bias was at work, one would expect judges who were aware of the contents of the trunk to rule in favor of the police at a higher rate, even though in both cases the legal issue was probable cause to search the car. But that’s not what happened: the responses of the foresight and hindsight groups were extremely similar. What explains this unusual result? One possibility is that the judges used System 2 deliberation, applying the intricate analysis demanded by the Fourth Amendment, to rule without regard to what was found in the warrantless search. This could be true because there is a large array of rules to apply and because “the intricacy of this area of law signals to the judge that intuition might be inconsistent with law and therefore they will need to think through the rules created by the appellate courts carefully.” This explanation is consistent with a study (using undergraduates and law students as subjects) finding that legal training reduced bias from stereotyping when the subjects had been taught legal rules that applied to the cases they were given, but did not reduce bias when the training covered more general standards.

It is important to note, though, that this experiment was based on stipulated facts, and is about application of well-established law to those facts. It does not speak to the inference drawing/fact finding process and whether it would be influenced by hindsight bias. But the reality of most pretrial disputes may well be that the facts aren’t agreed. For example, in a motion to suppress in a criminal case such as this one, if the question is whether to believe an officer’s testimony that a defendant made a furtive motion or consented to a search, judges’ decisions about whom to believe

---

155 Id. at 28. In contrast, a different test of hindsight bias showed that judges were influenced by knowledge of the outcome of an appeal when asked to identify the most likely ex ante ruling on a Rule 11 motion for sanctions—a decision granting the judge far wider discretion, governed by more general standards. Guthrie, Blinking, supra note xx at 25–26. Numerous other studies have shown judges to be influenced by hindsight bias in other contexts. See generally Wistrich, Can Judges, supra note xx at 1316 (“The vast literature on the hindsight bias includes virtually no studies that fail to uncover evidence of the hindsight bias in ex post assessments of ex ante probabilities.”)


could well be influenced by what they know about the fruits of the search.158

Another possibility is that given the frequency of probable cause hearings, judges develop rule of thumb heuristics applicable to common fact patterns in order to guide their rulings. They might, for example, refuse to find probable cause based only on vague assertions that the police smelled drugs.159 In either case, this study provides evidence that some circumstances—particularly well-developed facts and intricate legal rules—can cue judges to use their professional analytical skills to trump the automatic heuristic response.

D. Implicit Bias

This section looks at heuristics’ more insidious potential effects; they are the gateway to the operation of implicit biases based on factors such as race, gender, national origin, age, and religion.160 “Implicit biases are discriminatory biases based on implicit attitudes or implicit stereotypes.”161 We say that these attitudes and stereotypes are implicit when the person holding them is not intentional or even conscious of the bias.162 They can be especially difficult for the law to deal with, because implicit biases may be different from a person’s conscious beliefs but nevertheless affect the person’s actions.

One very common way to try to measure degrees of implicit associations between stereotypes and individual images is the “Implicit Association Test” (IAT), which is administered by Harvard University163 and has been taken by millions of people over the years.164 The IAT is a sorting

---

158 In am indebted to Jeremy Fogel, former U.S. District Judge, former Director of the Federal Judicial Center, and current Executive Director of the Berkeley Judicial Institute, for this point and example.

159 Wistrich, *Can Judges*, supra note xx at 1318. RGW considered both these possibilities and noted that the issue requires further study.


161 Anthony G. Greenwald & Linda Hamilton Krieger, *Implicit Bias: Scientific Foundations*, 94 CAL. L. REV. 945, 951 (2006). In social psychology, an “attitude” is “an evaluative disposition that is, the tendency to like or dislike, or to act favorably or unfavorably toward, someone or something.” *Id.* at 948. A “stereotype” is “a mental association between a social group or category and a trait.” *Id.* at 949. For example, a stereotype might be an association between old people and driving skill, and the attitude (unfavorable) would be that older people are poor drivers.

162 *Id.* at 946.


164 For a fuller discussion of the IAT, see Anthony G. Greenwald et al., *Measuring Individual Differences in Implicit Cognition: The Implicit Association Test*, 74 J.
test in which the subject is asked to match categories and traits that are often associated with implicit biases – such as Male/Female sorted with Career/Family. It tends to take longer to tap the computer key that matches the categories with traits that are contrary to stereotypes. For example, it may take longer to match words or images associated with Female to those associated with Career than to those related to Family. Those time differences generate a score that reflects how strong an association the IAT-taker has between the category and the trait. A person taking the test described, for example, might learn that the test has shown them to have a mild association between Female and Family. The best known IAT focuses on race, but there are also tests looking at associations based on age, religion, political party, sexual orientation, mental illness, weight, and more.

The collective results of years of IAT tests is worrying:

In the case of race, scientists have found that most European Americans who have taken the test are faster at pairing a White face with a good word (e.g., honest) and a Black face with a bad word (e.g., violent) than the other way around. For African Americans, approximately a third show a preference for African Americans, a third show a preference for European Americans, and a third show no preference.166

Showing an unconscious association is not the same as showing that people act on that association, and research on the latter is ongoing. There is, however, evidence that implicit associations can affect decisions, including hiring,167 medical treatment,168 and perceptions of a suspect’s dangerousness.169 There are also studies that do not show a relationship between IAT scores and behavior.

165 For a detailed description of the tasks involved in this version of the IAT, see Greenwald & Krieger, supra note xx at 952.
There have been few experimental studies trying to assess judges’ implicit bias. The RGW trio did one test of black/white bias among generalist judges in criminal cases, using state court judges (133 state or local trial judges) as subjects. Their IAT results showed that judges, like most white Americans, more closely associate African Americans than whites with negative concepts. The results of their experiments showed that those biases sometimes influenced judgments in hypothetical cases. It also showed, though, that in some instances (when the judges were aware of a need to monitor their own responses for the influence of implicit racial bias and motivated to suppress it) judges could focus more consciously on the issue of race and compensate for their implicit bias.

Two of their tests did not involve explicit mention of race, but were preceded by a subconscious “priming” task in which the test subjects were (or were not) exposed to words associated with black Americans. They found in both cases that “judges who expressed a white preference on the IAT were somewhat more likely to impose harsher penalties when primed with black-associated words than when primed with neutral words, while judges who expressed a black preference on the IAT reacted in an opposite fashion to the priming conditions.”

On the other hand, when race was more explicit, there was no significant correlation between conviction and the judges’ IAT scores. “Judges who exhibited strong white preferences on the IAT did not judge the white and black defendants differently, and neither did judges who expressed black preferences on the IAT.” The researchers attributed this result to the judges’ awareness (even though not told, most figured it out) that they were being studied for racial bias, and they were highly motivated to avoid being seen as biased in carrying out their assigned task.

---

171 Id. at 1197.
172 Id. at 1212 – 13.
173 Id. at 1217.
174 Id. at 1219. These results were quite different from results using the same materials as this third test by Professors Sommers & Ellsworth, who had only white subjects and who found that 90% of the participants who read about an African-American defendant said that they would convict as compared to only 70% of participants who read about a Caucasian participant. Samuel R. Sommers & Phoebe C. Ellsworth, White Juror Bias: An Investigation of Prejudice Against Black Defendants in the American Courtroom, 7 Psychol. Pub. Pol’y & L. 201, 216 – 17 (2001).
175 Rachlinski, Does Unconscious, supra note xx at 1223 – 24.
Absent the triggering and motivation, there are grounds to worry that judges are not sufficiently self-aware to correct for bias. The same researchers asked a group of judges to rate their ability to “avoid racial prejudice in decisionmaking” relative to other judges attending the same conference. Thirty five out of thirty six (97%) of the judges rated themselves in the top half, and eighteen out of the thirty six (50%) rated themselves in the top quarter of ability. The RGW group noted, “We worry that this result means that judges are overconfident in their ability to avoid the influence of race and hence fail to engage in corrective processes on all occasions.”

As part of a larger experiment designed to see whether specialist judges would be more able to overcome heuristics, the RGW group tested whether the race of a debtor would affect the decision of a bankruptcy judge. They gave the judges an elaborate fact pattern involving discharge of student debt for undue hardship; the test for discharge required the judge to make findings about three specific factors. Everything was identical, except that the name of the debtor varied, using names that prior researchers had indicated were associated with African-Americans or with Anglos. The results showed no statistically significant racial bias. Judges given debtors with African-American sounding names discharged a mean of 56.2% of the debt ($47,106), while those given debtors with Anglo-sounding names discharged a mean of 57.9% ($48,506) of the debt. Although not correlated with race, it is interesting to note that Democratic judges discharged a mean of $50,972, while Republican judges discharged a mean of $34,232. This would seem to indicate that there was enough leeway involved in the decision

---

176 Id. at 1225. This result is consistent with a study of administrative law judges: 97.2% placed themselves in the top half compared to other attendees with respect to their capacity for avoiding bias in judging (50% in the top quartile and 47.2% in the second quartile). Chris Guthrie et al., The “Hidden” Judiciary: An Empirical Examination of Executive Branch Justice, 58 DUKE L.J. 1477, 1519 (2009).

177 Id. at 1226.

178 Rachlinski, Bankruptcy Judge, supra note xx at 1245.

179 The names identified as African-American sounding were Ebony, Latonya, Kenya, Latoya, Tanisha, Lakisha, Tamika, Keisha, and Aisha. The names identified as Anglo-sounding were Kristen, Carrie, Laurie, Meredith, Sarah, Allison, Jill, Anne, and Emily. They were used as names on fake resumes on a test of employment bias in Marianne Bertrand & Sendhil Mullainathan, Are Emily and Greg More Employable than Lakeisha and Jamal? A Field Experiment in Labor Market Discrimination, 94 AMER. ECON. REV. 991, 991-92 (2004).

180 Rachlinski, Bankruptcy Judge, supra note xx at 1247. Some of the judges reported after the presentation of the test results that they did not remember the name of the debtors in their cases, perhaps because their attention was elsewhere. The factual and legal complexity of the task likely triggered the use of System 2 cognition.

181 Id. at 1247 - 48.
to generate differences – but for this group of bankruptcy judges, political attitudes about debtors and government assistance were far more significant than race in influencing the judges’ decisions.\textsuperscript{182}

In observational studies examining the potential impact of judges’ personal characteristics, including race and gender, patterns emerge in cases in which that characteristic might be salient, indicating that the judges’ “experience” differs in ways that can affect outcomes. It would make sense that a judge’s lived experience as a member of a traditionally disadvantaged group might affect the inferences that the judge would draw. Many of the studies are in the context of the criminal justice system. On the civil side, one context is especially enlightening: cases alleging employment discrimination based on race or gender. For example, studies found that:

- Female judges favored plaintiffs in employment discrimination cases relative to their male counterparts.\textsuperscript{183}
- For female trial judges, the average predicted probability of ruling for the sex discrimination plaintiff was 35%. For male judges, this number was only 20%.\textsuperscript{184}
- There was a 126% increase in the likelihood of a black trial judge ruling in favor of the EEOC’s race discrimination claim compared to a white trial judge.\textsuperscript{185}
- Plaintiffs in racial harassment cases were successful in 46% of their cases before black judges but less than half as often before white judges; logistic regression analysis indicated that on average, plaintiffs before black judges were 3.3 times more likely to win than before white judges.\textsuperscript{186}

\textsuperscript{182} A different group of experimenters examined implicit bias targeting different groups, testing for implicit bias regarding Asian-Americans and Jews. They administered the test to 180 federal judges and 59 state judges. On their IAT tests, the judges showed moderate to strong implicit bias against both Asians (relative to Caucasians) and Jews (relative to Christians). Justin D. Levinson et al., \textit{Judging Implicit Bias: A National Empirical Study of Judicial Stereotypes}, 69 FLA. L. REV. 1, 35 (2017).


\textsuperscript{185} \textit{Id.} at 794. Black judges were also more likely than white judges to rule in favor of plaintiffs in sex discrimination cases, \textit{Id}.

One recent study is intriguing, supporting the possibility that the judge’s experience, rather than sheer identity, can affect inferences. Using a data set on judges’ families and a data set of nearly 1,000 gender related cases, Professors Glynn and Sen found that judges with at least one daughter vote in a more liberal fashion on gender issues than judges with no daughter. “The effect is robust, and appears driven largely by male Republican appointees.”187 The most likely explanation: “having daughters leads judges to learn about issues that they ordinarily would not be exposed to – such as discrimination on the basis of pregnancy, Title IX, and reproductive rights issues.”188 In other words, having daughters can change the judges’ experience and view of the world in ways that makes inferences in favor of female litigants more plausible.

Federal judges in particular live in a somewhat rarified, elite environment in which they are isolated from both lawyers and the public, and they are treated with great deference. All of that can be a good thing when it comes to avoiding ethical conflicts and respecting the role of the courts, but it makes it even more challenging for them to have experiences that give them insights into the personal and professional experiences of everyday litigants. The judges’ own experiences will predictably affect the generalizations they bring to inference-drawing, and their heuristics can amplify the biases that that result.

E. Cultural Cognition

Another type of non-merits influence that can affect the ways that judges draw inferences is cultural cognition. “Cultural cognition refers to the tendency of individuals to conform their beliefs about disputed matters of fact . . . to values that define their cultural identities.”189 Cultural cognition does not grow solely out of an over-use of heuristics. “On the contrary, multiple studies have found that the individuals most proficient in and most disposed to resort to System 2 modes of information processing are even more likely to construe information in a manner that evinces identity-protective

187 Adam N. Glynn & Maya Sen, Identifying Judicial Empathy, 59 AM. J. POL. SCI. 37, 38 (2015). Having daughters had no effect on cases without a gender dimension. Id. at 47.
188 Id. at 51. Cf. Patricia Yancey Martin et al., Gender Bias and Feminist Consciousness Among Judges and Attorneys: A Standpoint Theory Analysis, 27 SIGNS: J. OF WOMEN IN CULTURAL AND SOCIETY 665, 667 (2002) (arguing that “individuals’ social location in the sex-gender system affects their experiences, interpretations, and, ultimately, consciousness within and beyond legal contexts.”).
reasoning." Nevertheless, a cluster of System 1 heuristics form the backbone of cultural cognition. The mechanisms that result in cultural cognition include the availability heuristic, biased assimilation of information, credibility (social affinities used to assess credibility and trustworthiness of sources of information), and affect heuristic (identity-protective motivations “to conform one’s beliefs to those of like-minded others in order to avoid dissonance and protect social standing”). In fact, much of the difference in perception “coheres with membership in groups integral to personal identity, such as race, gender, political party membership, and religious affiliation.” The net result of these factors can be that perception of facts is “unconsciously motivated by a person’s identity-defining commitments.” The theory of cultural cognition also complements the heuristic model by “showing how one and the same heuristic process (whether availability, credibility, loss aversion, or affect) can generate different perceptions of risk in people with opposing outlooks.”

Researchers at Yale Law School’s Cultural Cognition Project have done two particularly noteworthy experiments to assess the impact of cultural cognition on legally relevant fact finding. Both studies examined the test subjects’ interpretations of what they saw on videotapes. One was an edited video that had been the subject of a U.S. Supreme Court opinion, one in which Justice Scalia’s majority saw no possibility of differing interpretations. The study, however, demonstrated that people with different cultural

---

192 “[C]ultural affinity and cultural difference supply the relevant in-group/out-group references that in turn determine whom people see as knowledgeable, honest, and unbiased, and thus worthy of being credited in debates about risk.” Id. at 512.
193 Dan Kahan et al., Cultural Cognition of Scientific Consensus, 14 J. RISK RESEARCH 147, 150 (2010). See also Kahan, Who Fears, at 506 (identifying biased assimilation and source credibility as the mechanisms behind cultural cognition).
194 Kahan, Who Fears at 505.
195 Kahan, They Saw at 860. See also Dan M. Kahan, The Politically Motivated Reasoning Paradigm, Part 1, in EMERGING TRENDS IN SOCIAL & BEHAVIORAL SCIENCES at x (2016) (comparing and contrasting politically motivated reasoning with confirmation bias; the goal of the former is identity protection).
196 Kahan, Who Fears at 509.
commitments drew very different inferences about danger and fault from watching the same videotape of a speeding driver fleeing a police car.\textsuperscript{197}

Another dramatic demonstration of the impact of cultural cognition involved a video of protesters in a public place.\textsuperscript{198} The test subjects\textsuperscript{199} were shown a video of a political demonstration, but half believed it to be a protest outside an abortion clinic\textsuperscript{200} and half believed that the demonstrators were protesting the military’s “don’t ask/don’t tell” policy outside of a recruitment facility.\textsuperscript{201} (Neither was actually true.) Whether the subjects thought the protesters obstructed and threatened pedestrians “[depended] critically on the relationship between the demonstrators’ causes and the subjects’ own values. . .

Responses on other items—such as whether the protesters ‘screamed in the face’ of pedestrians and whether the protesters intended only to persuade or instead to threaten—displayed similar patterns.”\textsuperscript{202} In other words, if the test subjects disagreed with the protesters, they were far more likely to infer obstruction, screaming, and intent to threaten; if they agreed with the protesters, they were less likely to infer those actions. And yet all of these groups watched the identical 3 ½ minute video.

There is little empirical testing of the effect of cultural cognition on judges. One study examined the reactions of a sample of judges to social science studies relating to the deterrent effect of the death penalty. Although the judges applied evidence law fairly neutrally in making decisions about admissibility, the judges’ pre-existing opinions and political outlooks correlated with their decision about whether to afford “dispositive weight” to

\textsuperscript{197} Dan M. Kahan et al., \textit{Whose Eyes Are You Going to Believe?: Scott v. Harris and the Perils of Cognitive Illiberalism}, 122 HARV. L. REV. 837 (2009). In the case that ended up at the Supreme Court as \textit{Scott v. Harris}, five judges (the trial judge, the Eleventh Circuit panel, and Justice Stevens) and eight judges (Justice Scalia and the Supreme Court majority) reached opposite conclusions about whether the only reasonable inference from the video was that the fleeing driver posed a significant danger to police or the public. Harris v. Coweta County, No. CIVA 3:01CV148 WBH, 2003 WL 25419527 (N.D. Ga. Sep. 25, 2003), rev’d in part, 433 F.3d 807 (11th Cir. 2005), rev’d sub nom. Scott v. Harris, 550 U.S. 372 (2007).

\textsuperscript{198} Kahan, \textit{They Saw, supra} note xx at 863.

\textsuperscript{199} The subjects were 202 American adults who were selected randomly from a stratified national sample by Polimetrix, Inc. \textit{See id.} at 869 for more information.

\textsuperscript{200} The abortion clinic version of the video is available here: \url{https://youtu.be/k8ru-FE2v_8} (last visited Jan. 22, 2019).

\textsuperscript{201} The military recruiting version of the video is available here: \url{https://youtu.be/X3P1ACpL53k} (last visited Jan. 22, 2019).

\textsuperscript{202} \textit{Id.} at 878. Relatively few study participants from any group reported seeing spitting or shoving, regardless of their cultural identity. The researchers believe this to be evidence that differing values only affect contexts in which there is legitimate room for interpretation. It also helps show that the participants were not responding in a \textit{consciously} biased way. \textit{Id.} at 883.
the studies in determining the constitutionality of the death penalty.\textsuperscript{203}

Political science literature is rife with studies showing a correlation between judges’ legal decisions and their political preferences.\textsuperscript{204} And there is little question that when a normative position is intrinsic to the judicial decision to be made, judges’ ideology will be reflected in their decisions. But what about decisions in which the preference-salient fact is not legally relevant?

One experimental study by Professors Kahan, Hoffman, and others looked at judges and cultural cognition; it concluded that non-intrinsic cultural commitments do not influence judges who are using their System 2 deliberative reasoning to interpret somewhat ambiguous statutory language in order to apply it to a set of facts.\textsuperscript{205} More specifically, the judges/test subjects considered two hypothetical cases. In one, defendants who would be culturally attractive to one group and unappealing to another were charged under a statute prohibiting “littering, disposing, or depositing any form of garbage, refuse, junk, or other debris” on the grounds of a national wildlife preserve.” Both defendants were leaving empty water jugs that they intended to use to help themselves or others. In the second hypothetical, the issue was whether a statute prohibiting “knowing” disclosure of confidential information would cover actions by a police officer that were done intentionally but without knowledge that the action was illegal. The reasons for disclosure were the variable, this time polarized around attitudes toward abortion.

Comparing 253 sitting state court judges to a control group of 800 members of the public (including lawyers and law students), the study found that judges’ legal training “enables judges of diverse cultural identities to converge on ‘correct’ answers to statutory interpretation questions that trigger polarization along identity-protective lines in the public.”\textsuperscript{206} The same results were observed in the lawyers. Note, though, that the test cases involved situations which suggested clear “correct” answers to a legally-

---


\textsuperscript{204} See, e.g., Lee Epstein et al., Ideology and the Study of Judicial Behavior, in IDEOLOGY, PSYCHOLO., & L. 717–23 (Jon Hanson ed., 2012) (surveying political science literature on ideology and judicial decision making).


\textsuperscript{206} Id. at 411-12.
trained person. Items placed in the wildlife reserve to be useful are unlikely to be “garbage, refuse, junk, or debris.” Even more so, the oft-invoked principle that “ignorance of the law is no excuse” may have made it seem quite obvious to lawyers that an intentional act was “knowing.” This result—the ability of a clear legal principle to override heuristic biases—is consistent with Professor Girvan’s study (on law students) showing that legal rules can override heuristic biases, while standards and balancing tests do not.\(^{207}\) It would be interesting to see a study in which the interpretation-application task was in a context which would seem more indeterminate even to legally trained test subjects. Would the judges’ cultural identities exert more influence under those conditions?

This study also does not purport to address the question of whether judges’ cultural cognition affects their decisions about disputed facts, including inferences. The RGW group, in writing about a study of judges’ susceptibility to the affect heuristic, predicted that the impact on fact finding might be even stronger than that on legal analysis. “We cannot say for sure, of course, but we believe that emotion exerts even greater influence on fact finding. Many believe that facts are more uncertain than law, and the greater the indeterminacy the greater the opportunity for extrinsic influences, such as affect, to intrude.”\(^{208}\)

F. Judicial Demographics

Heuristics are mental processes that operate to foreground information, but they will bring forward different information for different people, because those people have different information and experiences (availability), and different attitudes (confirmation and affect heuristics, implicit biases, and cultural identities). These differences will lead to different generalizations and different inferences. Across the federal judiciary, the collective impact of heuristics would be somewhat less worrying if the judiciary were well-balanced in terms of professional and life experiences and cultural commitments. But is this the case?

The Congressional Research Service did a study of the demographics

---

\(^{207}\) Erik J. Girvan, Wise Restraints?: Learning Legal Rules, Not Standards, Reduces the Effects of Stereotypes in Legal Decision-Making, 22 PSYCHOL., PUBLIC POL’Y, & L. 31 (2016). On the other hand, this study contrasts interestingly with the RGW study of the impact of affect, finding that judges’ interpretation of an ambiguous statute varied with the sympathy one would feel with the defendant. See Wistrich, Heart, supra note xx at 877-88 and text accompanying notes xx, supra.

\(^{208}\) Id. at 904.
of the federal district and circuit courts as of June, 2017. Focusing on the district courts, where the bulk of the pretrial inferences are made, the data show a group that is heavily white, male, and older. Of the 570 active district judges at that time, only 194 (34%) were women. In terms of race, 406 (71%) were white, 81 (14%) were African American, 58 (10%) were Hispanic, 16 (3%) were Asian American, 1 (.2%) was America Indian, and 8 (1.4%) were multiracial. Of the ten judicial districts with the highest percentage of the population that is African American, four had no currently serving African American judges (the Southern and Middle Districts of Alabama, Southern District of Georgia, and Western District of Louisiana).

It is likely that the percentage of white and male judges has increased somewhat under President Trump. As of 2/9/19, of his 53 appointees to the federal district courts, he appointed 37 (69.8%) white male and 12 (22.6%) white female district court judges (so 92.4% of his appointees are white). He appointed 4 (7.5%) district court judges of either sex (2 male, 2 female) who are non-white.

Some research indicates that the presence of nontraditional judges may help to broaden the perspective of other judges. One type of study focuses on the courts of appeal and looks at “panel effects,” while others examine the possibility that a critical mass of nontraditional judges can make

---

209 Barry J. McMillion, U.S. Circuit and District Court Judges: Profile of Select Characteristics, Congressional Research Service (2017). While the study also contained information about sitting judges’ positions immediately prior to being appointed to their current federal judicial position, its categories are too broad to be very informative (e.g. “private practice”) and it does not contain information about all of the judge’s pre-bench professional work. The percentage of judges coming to the district bench directly from private practice has decreased over time. See Russell Wheeler, Changing Backgrounds of U.S. District Judges: Likely Causes and Possible Implications, 93 JUDICATURE 140 (2010).

210 Id. at 15.

211 Id. at 17. When gender and race are combined, the data show that as of June, 2017, 49.3% of federal district judges were white men, 21.9% were white women, 8.1% were African American men, 6.1% were African American women, 6.5% were Hispanic men, 3.7% were Hispanic women, 1.6% were Asian American men, and 1.2% were Asian American women. The report noted that all categories other than white men are considered to be “nontraditional” and that 56.1% of the active nontraditional district judges had been appointed by President Obama. Id. at 18-19. Statistics on age show that the federal district judges skew older. In June of 2017, the average age of a U.S. District Judge was 60.8 years (the median was 61.3). The largest group (269 judges—or 47.2%) are between the ages of 60 and 69. An additional 53 judges (9.3%) are 70 or older, making 56.5% of all federal district judges over 60 years of age.

212 Calculations were made using the search tool at Federal Judicial Center, Biographical Directory of Article III Federal Judges 1789-Present, available at https://www.fjc.gov/history/judges (last visited Feb. 9, 2019).
a difference. In that regard, the federal district courts are a mixed bag. In June of 2017, there was at least one active female district judge in 80 of 91 district courts, but there were no women serving on 11 district courts.

Thirty-seven district courts had only one active female judge. Lack of racial minorities was even more dramatic. In June of 2017, “there were African American judges serving on 44, or 48%, of the nation’s 91 U.S. district courts; Hispanic judges serving on 24 (26%) of the courts; and Asian American judges serving on 12 (13%) of the courts.” Twenty-three of the district courts that did have an African American judge had only one. Only seven courts have at least one active district judge from all three of the groups counted (“i.e., there was as at least one active African American, Hispanic, and Asian American judge serving on the court”).

From an income perspective, federal district judges are by virtue of their salaries alone in a place that makes them different from many litigants. For 2017, the annual salary of federal district judges was $205,100. The U.S. Census bureau puts the 2017 median household income at $61,372, so

---

213 See, e.g., Christina L. Boyd, Lee Epstein & Andrew D. Martin, Untangling the Causal Effects of Sex on Judging, 54 Am. J. Pol. Sci. 389, 406 (2010) (“For males at relatively average levels of ideology, the likelihood of a liberal, pro-plaintiff vote increases by almost 85% when sitting with a female judge.”); Pat K. Chew, Comparing the Effects of Judges’ Gender and Arbitrators’ Gender in Sex Discrimination Cases and Why it Matters, 32 OHIO ST. J. ON DISP. RESOL. 195, 200 (2017) (“Evidence indicates that when groups are comprised of more diverse members, those members learn from each other and provide checks on the correctness of shared information, ultimately leading to more accurate decisionmaking.”); Paul M. Collins, Jr., Kenneth L. Manning & Robert A. Carp, Gender, Critical Mass, and Judicial Decision Making, 32 L. & Pol’y 260, 265 (2010) (“Prior research suggests that the mere presence of decision makers from an underrepresented group in the overall decision-making environment may be enough to have a discernable effect on the output of that environment, so long as a certain ‘critical mass’ is reached.”); Sean Farhang & Gregory Wawro, Institutional Dynamics on the U.S. Court of Appeals, 20 J.L. ECON. & ORGANIZATION 299 (2004) (institutional norm of unanimity encourages deliberation, which in turn may allow numerical minorities on panels to influence case outcomes). But cf. Rachlinski, Racial Bias, supra note xx at 1227 (noting that white judges from eastern jurisdiction in study showed strong set of implicit biases even though the jurisdiction consisted of roughly half white and half black judges. “Exposure to a group of esteemed black colleagues apparently is not enough to counteract the societal influences that lead to implicit biases.”)

214 McMillion, supra note xx at 15.

215 Id. at 16.

216 Id. at 17.

217 Id.

judges are well above the median. In fact, a recent Economic Policy Institute study shows that the district judge salary is comfortably in the average salary of the top 10% of earners for 2017.

Taking all of this demographic information together, it seems possible that the experience of the federal district judges as a group may be skewed toward the experiences of the affluent, white, male majority. While studies do not seem to show across-the-board differences between the decisions of male and female judges, or between white judges and judges of color, there is evidence of race and gender differences when race or gender is a salient issue in the case. Further studies, examining more fine-tuned heuristics, implicit biases, and cultural commitments may find additional effects. In any case, conscious efforts to appoint a body of trial-level judges who are diverse across a number of measures would help to avoid overall tipping of inferences and even substantive law in favor of the experiences of a homogeneous group.

G. Summary

Where does all this get us? With respect to judges as a group:

- During the pretrial period, they make countless decisions that are based on the evaluation and drawing of inferences;
- Those inferences can differ significantly from judge to judge;
- Those differences often turn on differences in the judges’ “experience and common sense” and the generalizations that those factors produce;
- That “experience and common sense” expresses itself through the powerful operation of mental shortcuts – heuristics and biases – that work automatically and at a subconscious level in a way that can

---

221 See text accompanying notes xx supra.
222 For example, it can be argued that the heuristics and biases of the dominant culture have caused the substantive law of employment discrimination to be shaped in a way that turns certain inferences (or rejection of inferences) into legal presumptions. See Nancy Gertner, Losers’ Rules, 122 YALE L.J. ONLINE 109(2012), http://yalelawjournal.org/forum/losers-rules (noting presumptions regarding “stray remarks,” “honest beliefs,” and the requirements for comparators).
influence even deliberate decisionmaking;

- At times the heuristics pave the way for implicit biases and culturally motivated biases to affect judicial decisions, even in judges who are consciously committed to an unbiased and impartial judicial system;
- The heuristics and biases can be particularly hard to tame, because of a cluster of person-centered heuristics that falsely lead people (including judges) to believe that they are more correct, their opinions more universally held, and their ability to avoid biases more exceptional than is in fact the case;
- Judges’ legal and professional training does not overcome the powerful influence of the unconscious processes in most cases that have been studied;
- In situations involving legal analysis (as opposed to fact-finding), judges have sometimes shown themselves able to overcome the influence of hindsight bias and cultural cognition to apply a clear, applicable rule, but there are as yet no studies with comparable results for fact-finding;
- In situations where the danger of racial bias is explicit, motivated judges may be able to consciously offset the bias when applying legal standards to a fact pattern. However, when racial issues are suggested unconsciously, implicit associations influence the judge.\(^{223}\)
- The demographics of the federal judiciary support a concern that the collective implementation of “experience and common sense” will skew both individual cases and the development of the law generally in the direction of the predominant judicial background, to the detriment of a wider array of experiences.

More studies will be needed for a richer understanding of judicial cognition. But there is enough information already to give reasons for concern, especially given the pervasive nature of judicial fact inferences. The next section will consider what systemic changes might help the inference process operate in a more reliable way.

VI. IMPROVING THE SYSTEM

Inferences are certainly not going away. But there are measures that the federal courts could take to lessen the impact of problematic heuristics on the pretrial process. First, sustained training on heuristics, biases, and ways to combat them might have some effect. Second, practices that activate

\(^{223}\) But see Rachlinski, *Bankruptcy*, supra note xx at 1246-48.
deliberative thinking and encourage consideration of competing inferences can be institutionalized. Third, certain rules of procedure (and their interpretations) should be changed in ways that (1) enrich the factual record on which the judges rule; and (2) discourage methods of analysis that allow the judge to compare inferences.

A. Judicial Education

Court systems have already begun to provide judges with education about the existence of heuristics and techniques to help deal with them, particularly in the context of implicit bias. The Federal Judicial Center has recently developed a set of judicial competencies, defined as “an area of proficiency essential for successful performance in a job.”224 The knowledge competencies include a broad array of types of social cognition.225 The competencies guide the judicial education curriculum, and the FJC has over the last several years offered numerous sessions on heuristics (including implicit bias), including it in new judge orientation, programs about “the art of judging” generally, and in substantive-law-specific programming.226 Similarly, the National Center for State Courts did a three-state pilot project to teach judges and court staff about implicit bias, and now has helpful resources on its website.227

Although it is important that the courts are doing this kind of education, there is reason to doubt that education alone can have a decisive impact, particularly if it is delivered in short, episodic programs. The extensive literature on de-biasing training is beyond the scope of this article, but two points are important to note here.

---

224 FEDERAL JUDICIAL CENTER, COMPETENCIES FOR UNITED STATES JUDGES v (June 2018).
225 The list includes cultural practices and norms inside and outside of the United States; social customs of groups besides one’s own; differences in communication styles among cultures; impact of implicit bias; impact of socio-economic status on personal decision making; different norms for interacting with authority figures; and experiences specific to race, religion, ethnicity, gender, sexual orientation, and gender identity. Id. at 8.
226 Telephone conversations with Judge Jeremy Fogel, then-Director of the Federal Judicial Center, educator Denise Neary, and researchers Jason Cantone and Beth Wiggins, Federal Judicial Center, September - October 2018.
First, some types of training appear to be more effective than others in the short run, and they are more effective at raising awareness than at spurring the adoption of behavioral countermeasures. For example, one 2014 study tested 17 methods for reducing implicit bias and found only 8 to be effective at reducing implicit preferences for whites as compared to blacks.\textsuperscript{228} Another meta-study found that the positive effects were “greater when training was complemented by other diversity initiatives, targeted to both awareness and skills development, and conducted over a significant period of time.”\textsuperscript{229} It is quite clear that merely telling judges to “do better” or “avoid bias” will not be effective.\textsuperscript{230}

Thinking more specifically about judicial education, it may be that the most effective types of training in using deliberative rather than intuitive processes might be a combination of informational programs about heuristics (so that judges recognize that it is a “thing”),\textsuperscript{231} coupled with suggested skills that promote conscious thought. These could include techniques like “consider the opposite” or “consider alternatives,”\textsuperscript{232} structured methods of calling mental attention to perspectives and experiences other than the judge’s own instinctive ones.\textsuperscript{233} Education designed to increase motivation to minimize the impact of subconscious biases might also enhance the impact of the programs. This might include “periodic reaffirmation of professional norms,” particularly in avoiding bias.\textsuperscript{234} More generally, “motivated self-
regulation” can result from awareness of bias that is inconsistent with a judge’s conscious commitments to fairness and accuracy. One study found:

the motivated self-regulation strategy operates by making people aware of their biased responses that stand in conflict with their personal beliefs. The resulting feelings of guilt and disappointment with the self then lead to a cascade of consequences that help people monitor for and regulate potentially biased responses in the future.235

Judicial motivation would likely be essential to successful education. “For any change to take place, it is imperative that judges be motivated to overcome their automatic processes.”236

Second, there is almost no evidence that anti-bias training has an effect in the long run. The same researchers who had found 8 out of 17 of the tested interventions against implicit bias to be effective did two more studies and reached a depressing conclusion: “In two studies with a total of 6,321 participants, all nine interventions immediately reduced implicit preferences. However, none were effective after a delay of several hours to several days.”237 On a more positive note, one study did show small effects of training two years later.238

Education about the impact of nonconscious thought processes, then, might be of some help, especially if done as part of overall institutional systems and repeated over time. It will be important for judicial education officials to attempt to assess the effectiveness of the programs they provide, both to identify the most promising techniques and to try to gauge the judiciary’s level of achievement of the desired competencies. This type of assessment is difficult, because self-reports about unconscious processes are apt to be unreliable (“No, I am not influenced by unconscious cognition.”)

236 Simon, Psychological Model, supra note xx at 140.
237 Calvin K. Lai et al., Reducing Implicit Racial Preferences II: Intervention Effectiveness Across Time, 145 J. EXPERIMENTAL PSYCHOLOGY: GEN’L 1001 (2016), available at DOI: 10.1037/xge0000179. See also Bezrukova, supra note xx at 1242 (“Comparing the immediate versus long-term effects of diversity training, we found that diversity training effects on reactions and attitudinal/affective learning decayed over time.”)
Education, though, will not provide the magic bullet, and so further structural changes will be important.

B. Deliberation-Prompting Practices

The types of educational programs described above are primarily ways to equip judges to move from the automatic System 1 responses into deliberate System 2 thinking. In addition to training designed to raise awareness of the need to do so and some techniques for activating diverse perspectives, there are ways to create institutional requirements that foster more effortful thought. They would also have the benefit of making decisions more transparent and demonstrating that even arguments that the court has not adopted have been heard. Some techniques operate at the individual-judge level, while others would require the creation of systemic structures.

Scholars have made a number of suggestions about practices that can prompt judges to make themselves aware of the impact of their own unconscious processes. A jurisdiction might even mandate or incentivize use of some of these behaviors, at least in certain types of cases. Some – such as “consider the opposite” – have been discussed above as a technique to help make the judge aware of the contestability of her own intuitions. A related technique is referred to as “perspective-taking.” Here the judge would adopt the viewpoint of other individuals and examine the scenario through their life experiences.239 The use of decision protocols such as checklists and scripts may also encourage deliberation.240 The checklists could even require the judge to indicate that he or she had used a deliberative technique such as “consider the opposite” in order to check for the potential impact of heuristics. These physical reminders of an intention to counteract the impact of heuristics can help the judge form a new habit of doing so.

Retrospective self-analysis could also be helpful. Judges could record decisional data and periodically review it to look for patterns.241 As Professor

---

239 R&W, Implicit, supra note xx at 112. Note, however, that this type of reasoning runs the risk of using stereotypes in making assumptions about what a person with a particular identity would be thinking. Id.

240 Id. at 119; NCSC, Addressing Implicit Bias in the Courts at 7; Wistrich, Head, supra note xx at 910. Although these can help, they are also easily circumvented through unthoughtful, rote compliance if a judge is not actually motivated to locate and reconsider the impact of heuristics and biases. Research showing that the application of clear rules can overcome heuristic biases might argue in favor of replacing flexible legal standards with more fixed rules, but attempting to do so could be difficult and, worse, substantively undesirable.

Kang and his co-authors noted, “Judges need to count.” A failure to live up to one’s own expectations can provide a judge with motivation to continue to try to consider others’ experiences and to think hard about the universality of “common sense.”

When a decision might be influenced by emotion or affect, the judge should be encouraged to engage with that emotion rather than to try to suppress it. “Although some judges profess to follow suppression strategies, there is no evidence that such strategies are effective. In general, trying not to think about something not only is ineffective but may even have an ironic rebound effect.” Similarly, judges should avoid making difficult decisions when tired, stressed, distracted, or pressured.

The capstone of a more slow and conscious process would be the practice of writing opinions that go into some detail about the reasons for the decision. Jurisdictions wanting to institutionalize this practice could, through statute or procedure rules, require opinion writing. Rethinking the way in which opinions are written could also promote deeper thought. Traditionally, once a judge has reached a conclusion, opinions tend to be written in a way that marshals every argument in support of that conclusion and downplays or ignores opposing arguments. Instead, opinions could recognize strong counterarguments – which some scholars argue can help judges avoid “decisions based on their own culturally biased motivations.”

---

242 Kang, Implicit Bias, supra note xx at 1178
244 NCSC, Addressing Implicit Bias in the Courts at 9. In one of the most disturbing studies on the impact of judges’ mental state on decisions, economists from Louisiana State University found that juvenile court judges who were LSU graduates handed down more severe sentences to black defendants during the week following an unexpected football loss. Ozkan Eren & Naci Mocan, Emotional Judges and Unlucky Juveniles, 10 AM. ECON. J.: APPLIED ECON. 171 (2018).
245 R&W, Implicit, supra note xx at 117 – 18.
246 See, e.g., FED. R. CIV. P 11(c)(6) (“An order imposing a sanction must describe the sanctioned conduct and explain the basis for the sanction.”); FED. R. CIV. P. 65(b)(2) (“Every temporary restraining order issued without notice must . . . describe the injury and state why it is irreparable.”).
247 Kang, Implicit Bias, supra note xx at 43-44; Simon, Psychological Model, supra note xx at 137.
248 Secunda, Cognitive Illiberalism, supra note xx at 391. See also Kang, Implicit Bias, supra note xx at 43-44; Kahan, They Saw, supra note xx 896 (suggesting that jurors and appellate panels might adopt a practice of “deliberate depolarization” in which a person expressing his or her own opinion would also express the strongest counterargument). This is easier said than done, however. The usual process of analyzing complex materials involves
This could both highlight the need to consider those counterarguments, and make the process and end result of the decision clearer both to the litigants and to members of the public who are interested in the dispute.

Deliberation is also more likely when the judge has adequate time to consider his or her decision.249 In addition to the usual admonitions about taking time to reflect, and avoiding making decisions when stressed or tired, one statutory reform could be useful here. A recent study has demonstrated that the so-called “Six Month List,” a law mandating the disclosure, twice a year, of each judge’s pending cases and motions, may actually have an effect on case outcomes. The forced deadlines (to avoid public shaming about perceived delay in disposing of cases) may in fact change the way judges rule.250

Other potential interventions create systems that allow judges to interact in ways that could help self-assess their decisions, look for patterns, and provide each other with peer feedback. A number of systems are possible, including collegial peer review committees that meet periodically to talk about their decisions and associated issues.251 Another alternative would be anonymous peer review, including concrete suggestions about how to improve performance.252

Systemic practices designed to incentivize or require judges to think about alternatives before reaching conclusions, to force themselves to go through each step of the analysis, and to articulate reasons for their decisions can be helpful in combatting misleading heuristics and in increasing transparency for the parties and the public. All of this also could help to reduce a bias blind spot held by the judges when faced with heuristics and training. But like training, they can do only so much. Our brains are

\[ \text{\textit{the judge, whose brain seeks coherence, unconsciously changing the evaluation of prior positions, and this process operates at an unconscious level. Simon, Psychological Model, supra note xx at 19.}} \]

\[ \text{\textit{249 See R&W, Implicit, supra note xx at 111 – 112.}} \]


\[ \text{\textit{251 R&W, Implicit, supra note at 18; Rachlinski, Does Unconscious Bias, supra note xx at 1230. For example, a small group of judges could meet periodically to talk about their rulings on summary judgment motions and the reasons for them.}} \]

\[ \text{\textit{252 Id. As the authors note, it seems unlikely that judges would find this option attractive. A related example, also with political drawbacks, would be to identify certain types of cases in which the subconscious influence of heuristics could be particularly troubling, and handle appeals from those orders differently, assigning a de novo standard of review and a diverse panel of judges. Rachlinski, Does Unconscious Bias, supra note xx at 1231.}} \]
programmed to minimize effortful thinking, and our socialization encourages us to reach decisions that are consistent with our political identities. Further changes are therefore necessary in order to enhance judicial fact-finding and decrease the instances in which the inferences of a single judge will terminate a case.

C. Procedure Rules

Procedural standards can do something that education and awareness cannot: they can structure rules so as foster the development of a fuller factual record, leaving fewer gaps to be filled by inferences and informing those inferences with a richer and more nuanced set of facts. They can also make choices that give great weight to the right to trial by jury and minimize the ability of a single judge’s inferences to end litigation.

1. Pleadings: Stop Comparing Inferences

In theory, judges are not finding facts when ruling on motions to dismiss on the pleadings. During the last few decades, however, changes in the application of the procedure rules have led to a practice in which judges allow cases to move forward toward discovery only if they find the nonmovant’s inferences to be at least as credible as the alternatives. That makes the judge’s assessment of inferences overly determinative, especially when one realizes that those inferences are not some kind of neutral mathematical calculation.

Since the Supreme Court’s decision in Ashcroft v. Iqbal, judges ruling on motions to dismiss for failure to state a claim have been tasked with taking only “facts” as true and accepting inferences only if they are “plausible.” Although the opinion disclaims a requirement that the inference must be “probable,” it rejects “mere possibility.” In fact, the Court has refused to find an inference to be plausible if it believes another one to be more likely. Iqbal characterizes the result in Twombly as a decision that the phone companies’ parallel conduct was “more likely explained by” an inference of free market behavior than illegal agreements. Similarly, the Iqbal

---

253 Through a kind of definitional sleight-of-hand, the judge’s decision that an inference cannot be supported is defined as a question of “law.” See Elizabeth Thornburg, Law, Facts, and Power, 114 PENN STATIM 1 (2009), at http://pennstatelawreview.org/114/114%20Penn%20Statim%201.pdf (last visited Dec. 11, 2018) (discussing the policy-based nature of the law/fact distinction).
plaintiffs’ inferences of discriminatory conduct were found implausible: “given more likely explanations, they do not plausibly establish this purpose.”

Comparing inferences in this way puts far too much weight on the judge’s own inference choices, colored by the judge’s experience and attitudes. The question should not be whether the pleaded inference is at least as probable (in the opinion of the judge) as others; it should be whether it is possible. The risk of heuristic bias is so unavoidable under current law, that any analysis of pleading sufficiency that turns on comparing inferences must be abandoned. We cannot know whether experiences and heuristics contributed to the different inferences of the white Alabama trial judge and black Florida appellate judge in Lewis v. Bentley, but a rule structured so that they were not called upon to compare the plausibility of racial discrimination and state-level control would decrease this concern.

Since Rule 8(a) itself does not mandate comparisons as part of the Iqbal-mandated plausibility analysis, this change in analytical method could be made without a rule amendment. Judicial practices, though, would need to change; judges should not dismiss cases because they believe an inference contrary to the plaintiff’s is more probable.

As to types of cases in which slanted heuristics are particularly likely, a solution would need to come from outside the Federal Rules of Civil Procedure. The trans-substantive nature of the rules precludes changes that single out specific substantive areas for different treatment. Outside of the rules, though, changes in substantive law that establish certain presumptions might help highlight issues in which inferences are particularly likely to be unreliable because of the limits of generalizations or social stereotypes.

2. Summary Judgment: Consider Only Inferences in Favor of Non-Movant

The analysis in summary judgment cases works similarly, but no change in the law would be necessary. Proper application of existing legal standards does not allow the judge to choose among reasonable inferences. Rule 56 instructs judges to grant summary judgment only if there is no “genuine dispute as to any material fact” and the “movant is entitled to

\[^{256}\text{Id. at 680, 681.}\]

\[^{257}\text{See Section IV(A)(1), supra.}\]

judgment as a matter of law.”259 The temptation to compare inferences comes from adjectives that modify the word “inference.” In both summary judgment and directed verdict contexts, judges are told to let the case continue if the nonmovant’s inference is reasonable. “Credibility determinations, the weighing of the evidence, and the drawing of legitimate inferences from the facts are jury functions, not those of a judge, whether he is ruling on a motion for summary judgment or for a directed verdict. The evidence of the nonmovant is to be believed, and all justifiable inferences are to be drawn in his favor.”260

The problem with these descriptions lies not in the requirement of reasonableness, but in some courts’ practices of 1) considering even impeached evidence in favor of the movant; 2) considering inferences in favor of the movant; and 3) judging reasonableness comparatively, so that an inference only counts as reasonable if, in the judge’s view, it is superior to others. Jurors, the ultimate fact finders, are the ones entrusted to compare beliefs in different possibilities.261 That is not the job of a judge ruling on a motion that will end the case before it goes to the jury.

Greater emphasis on what may and may not be considered would help clarify the mental process that a judge ruling on a summary judgment motion should use. The judge needs to be acutely aware that she should consider all evidence in favor of the nonmovant, and only evidence in favor of the movant that has not been contradicted or impeached.262 (If evidence has been contradicted or impeached, the judge would be making a credibility decision, and that is not proper in the summary judgment context.) Note also that, in the case of movants’ evidence, this is limited to testimony from “disinterested witnesses.”263

In addition to the limits on what evidence to consider, proper application of Rule 56 requires that the judge consider only the inference in support of the nonmovant without comparing it to alternatives. If that inference is reasonable, considered in isolation, then the case should continue to jury trial. Unfortunately, some cases have considered pro-movant inferences in deciding that the nonmovant’s inference could not be found by

261 Clermont, Standards of Proof, supra note xx at 22.
262 Wilkerson v. McCarthy, 336 U.S. 53, 57 (1949) (directed verdict); Reeves v. Sanderson Plumbing Products, Inc., 530 U.S. 133, 151 (2000). Although Reeves is about post-trial motions, the opinion notes that the standard to be applied should be the same for summary judgment.
263 Reeves, 530 U.S. at 151.
a reasonable jury. For example, in the Chadwick case discussed in Section IV, the trial judge granted summary judgment because he rejected as a matter of law an inference that denying the plaintiff a promotion was motivated by her status as a woman with young children; to him, the inference that the same decision would have been made about a man with young children had not been sufficiently ruled out.

Other cases, especially those involving state of mind, reflect the same kind of failure to analyze the record in light of the limitations on summary judgment. In Jennings v. University of North Carolina, the trial judge, original majority opinion, and dissenting opinion on en banc review considered (and preferred) inferences in favor of the movant. The plaintiff had been a member of UNC’s women’s soccer team, and she alleged that the coach “persistently and openly pried into and discussed the sex lives of his players and made sexually charged comments, thereby creating a hostile environment in the women’s soccer program.” The Fourth Circuit’s en banc opinion concluded that the plaintiff had produced sufficient summary judgment evidence to merit consideration by a jury.

The judges who would have granted UNC’s motion for summary judgment considered contradicted and impeached evidence in favor of the movant, and they also relied on inferences in favor of the movant in finding the plaintiff’s requested inferences unreasonable as a matter of law. For example, they would have inferred that the coach’s frequent comments were lighthearted jokes, that the question “who are you fucking” showed concern that the plaintiff’s social life might be hurting her academic performance, and that the plaintiff’s distress at being cut from the soccer team showed that the impact of the coach’s behavior on her was not severe. Those pro-movant inferences were quite likely influenced by the judges’ understanding of the nature of athletics and player-coach relationships.

The Supreme Court itself was guilty of considering pro-movant

---

264 482 F.3d 686n (4th Cir. 2007) (en banc).
265 Id. at 691.
266 See, e.g., 482 F.3d at 719 (“simple teasing”) (Niemeyer, J., dissenting).
267 Id. at 720 (“[I]t was, in context, obviously an inquiry about what was occupying Jennings’ time.”)
268 Id. at 723.
269 Id. at 724 (“If Dorrance was then the coach of the men’s soccer team, he would just as surely have teased his male players about their weekends with their girlfriends as he lightly teased Jennings about her weekend with her boyfriend. Such teasing about a player’s social life is the norm on any collegiate athletic team, whether male or female.”)
One fact issue required a decision about whether a fleeing driver was endangering the lives of others. The summary judgment record in that case included deposition testimony as well as a videotape of a police chase. The decision turned not on facts depicted on the video (e.g. “was it dark outside?”) but on the inferences that could or could not be drawn from its depiction of the chase. The trial court and Eleventh Circuit opinions concluded that a reasonable jury could infer facts and reach conclusions in favor of the driver: he posed no threat prior to the chase, and during the chase he posed little or no threat to pedestrians or other motorists (the roads were largely empty and he remained in control of his vehicle). In contrast, Justice Scalia’s majority opinion for the Supreme Court rejected any such inferences as unreasonable as a matter of law, instead drawing the inference in favor of the police officer who rammed the fleeing driver’s vehicle. “Far from being the cautious and controlled driver the lower court depicts, what we see on the video more closely resembles a Hollywood-style car chase of the most frightening sort, placing police officers and innocent bystanders alike at great risk of serious injury.” Given his interpretation of the events depicted on the video, Justice Scalia chose the inference suggested by the movant police officer – the fleeing driver was endangering officers and members of the public – and therefore the nonmovant’s contrary inference was not reasonable. One can hope that this type of analysis is limited to cases involving videotape, as it does not purport to change the general rules of summary judgment analysis.

Improved awareness of proper summary judgment analysis could help avoid judges’ reliance on their own personal inferences to take cases away from juries. We would not need to wonder about the impact of differences in the experience of working mothers and working fathers (and societal attitudes toward both) on the trial and appellate judges in Chadwick. We would not have to worry that different experiences with team sports, coaches, and “locker room talk” would lead to different inference preferences.

---

271 Harris v. Coweta County, 433 F.3d 807, 819 (11th Cir. 2005).
272 550 U.S. at 379-80.
273 Justice Stevens’ dissenting opinion calls the majority on its improper summary judgment analysis, and also suggests that the Justices’ experience may have influenced their inferences. “I can only conclude that my colleagues were unduly frightened by two or three images on the tape that looked like bursts of lightning or explosions, but were in fact merely the headlights of vehicles zooming by in the opposite lane. Had they learned to drive when most high-speed driving took place on two-lane roads rather than on superhighways—when split-second judgments about the risk of passing a slowpoke in the face of oncoming traffic were routine—they might well have reacted to the videotape more dispassionately.” Id. at 390 n.1.
where the judge’s choice controls. Similarly, a plaintiff’s access to a jury trial would not be heavily influenced by whether a judge grew up driving on two-lane roads.

Important policies underlie Rule 56’s requirement that judges look at the evidence and inferences in the light most favorable to the nonmovant. The result of a denial of summary judgment is a case that proceeds to a jury determination. In addition to its Seventh Amendment significance, decisions by a multi-member jury may be better able to overcome the heuristics and biases discussed in this article. Its broader array of experiences and “common sense” make this possible. “[T]he controlled process of trial and the forced deliberation of lay jurors could provide a powerful antidote to not only overt biases and prejudices but also to the more subtle warping of perception and rational thought that stems from cognitive illiberalism and its cousins.”

When restricting analysis to the proper parts of the record, and to inferences in favor of the nonmovant, a judge’s role with respect to inferences is still very important, but its function is more narrowly defined. It also has the advantage of keeping the focus outside of the judge’s own personal probability assessment, reminding the judge that other people with different experiences can draw different but reasonable inferences. This approach might in some cases lead to increased costs to the litigants and the court because the litigation will continue, and the cost is more worrisome when the movant ultimately prevails. However, these costs must be balanced against potential advantages of avoiding the dismissal of meritorious claims, the costs of successful appeals, the importance of access to the courts, and the constitutional role of the jury as fact finder.

3. Develop a Record that Supports a Richer Narrative

a. Pleadings

Since the end of the twentieth century, the trend in civil procedure has been to worry about costs and efficiency at the expense of other values. As part of this trend, rule amendments and changes to interpretation of existing rules have restricted available procedures. Among other things, dismissals on the pleading and the granting of summary judgment have been made easier and the scope of discovery has been limited. All of these forces have resulted in decisions based on a sparser factual record – what one scholar has called

---

“narrative-erasing procedure.”275 The function of the civil justice system depends on competing narratives, and narrative richness helps decisionmakers choose from a fuller array of possibilities. “Because as a group judges are highly educated and very often members of a political or social elite, they may not share the same ‘stock stories’ with the litigants from outsider groups who appear before them. But we know narratives can create empathy and persuade in a way that cold logic or impersonal data cannot.”276

More information can lead to better inferences and more evenhanded results. Plaintiffs who do have access to a richer set of facts, then, should realize their narrative potential and plead the underlying story.277 Providing a more complete and full-bodied set of narratives can help judges combat automatic heuristic responses. For example, judges considering plausibility of pleadings may, if not given more information, rely on their own the narratives – those that their heuristics generate. “Given the privileged position of judges, who tend to be well educated and politically elite, [they may] unconsciously apply a master-narrative that differs radically from the narratives that . . . members of marginalized groups may wish to tell.”278 In order to increase the probability that the judge will be able to make decisions on a more information-rich record, procedure rules should be made and interpreted to discourage dismissal on the pleadings without an opportunity to do discovery.

Inferences are gap-fillers, and a record containing a more complete collection of circumstantial evidence can better inform the inference-drawing process. “The more gap filling and inferential thinking that a judge has to engage in, the more room there may be for explicit and implicit biases to structure the judge’s assessment in the absence of a well-developed evidentiary record.”279 Some of the additional information can provide the judge with facts that fall outside his or her experience, and it may even alert the judge to perspectives other than his or her own. The larger set of information inputs can also cue System 2 effortful thinking, increasing the

275 Anne E. Ralph, Narrative-Erasing Procedure, 18 Nev. L.J. 573 (2018). By “narrative,” she means “a particular representation of a series of events: a text or other embodiment of a certain telling or treatment of a story’s events.” Id. at 577. A narrative is not the same thing as a “story.” It is, rather, the telling of the events from some particular perspective.
276 Id. at 612.
278 Id.
279 Kang, supra note xx at 1163.
chances that an inference will be deliberate rather than automatic.\textsuperscript{280}

b. Discovery

Discovery decisions are the other significant procedural context that could improve inferences by providing a factually more complete record. Inferences come into the discovery picture because much of the information sought may be circumstantial rather than direct evidence. A party seeking discovery may be asking the court to order production of material that arguably supports an inference. For example, in a sex discrimination case involving John Doe’s intent, the discovery sought might be not “I heard John Doe say that he fired the plaintiff because she was female” but “John Doe regularly tells sexist jokes to male employees.” As was true in the cases described in Sections III and IV, the discovery relevance decision turns on whether the judge believes an inference from circumstantial evidence to the fact of consequence to the action is sufficiently significant to satisfy Rule 26(b).

In order to encourage discovery that will provide the court and the parties with a more complete narrative, judges drawing inferences can keep in mind the way the procedure and evidence rules talk about relevance.\textsuperscript{281} “Evidence is relevant if it (a) has any tendency to make a fact more or less probable than it would be without the evidence; and (b) the fact is of consequence in determining the action.”\textsuperscript{282} The language of the evidence rule makes it clear that the test should not require the discovering party to convince the judge that the requested information is sufficient to support an inference that is more probable than not. Rather, the information to be discovered is relevant if it makes the inference more probable than it would be without the information. The rule as written, then, encourages the judge to allow discovery of a broader set of information, and that information in turn can better inform the inference to be made on the merits.

Judges considering discovery requests should be aware that fact investigation, particularly when it is done by those with little initial information, may take the form of abductive reasoning. “Abductive reasoning” can be defined as “a creative process of using known data to

\textsuperscript{280} Cf. Guthrie, Blinking, supra note xx at 16 (judges did better on a more difficult CRT problem because more difficult problems suggest “to the test taker that reliance on intuition might be unwise.”)

\textsuperscript{281} Fed. R. Civ. P. 26(b)(1).

\textsuperscript{282} Fed. R. Evid. 401.
generate hypotheses to be tested by further investigation.” From the information available at a particular point in time, the discovering party may generate a hypothesis. If that hypothesis is correct, then the information sought should exist, and the discovery request is aimed at finding it. Rather than rejecting such requests out of hand as speculative or irrelevant, judges should entertain the possibility that inferential links between the information at hand, the hypothesis, and the information sought may be convincing.

The proportionality factors emphasized by the 2015 amendments to the discovery rules can also be problematic with respect to the impact of judges’ own experiences and biases. It is unlikely that they will be removed from the rule, but their application can be done with care, and with attention to the possible impact of non-universal generalizations. When the rule was adopted, the Advisory Committee left the question of burdens of proof somewhat ambiguous. The Advisory Committee note, though, provides both that “the [rule] change does not place on the party seeking discovery the burden of addressing all proportionality considerations” and that “party claiming undue burden or expense ordinarily has far better information – perhaps the only information – with respect to that part of the determination.”

Judges considering proportionality issues can avoid reliance on generalizations growing out of their own pre-bench practice experience by requiring parties to produce evidence to support the inferences relating to factors that are objective and measurable. Similarly, judges should be acutely aware of the possibility of generalizations that differ from those that they access most readily. Finally, for those proportionality factors that are inherently normative, judges should consciously summon and consider competing policy values.

Rule 26(b)(1) should be interpreted and applied with a policy-slant in favor of production, and this may in some cases increase costs for the producing party. It is nevertheless an important analytical starting point. Any increased out-of-pocket costs need to be considered in light of the benefits of further discovery that greater accuracy in decisionmaking could bring to private litigation’s function as an enforcer of legal norms.

---

283 ANDERSON, supra note xx at 379.
284 Id. at 57.
285 Advisory Committee Notes, 2015 amendments to FED. R. CIV. P. 26(b)(1).
286 See Jonah B. Gelbach & Bruce H. Kobayashi, The Law and Economics of Proportionality in Discovery, 50 GA. L. REV. 1093, 1111 (2016) (“the enumerated proportionality factors will inevitably involve unavoidable and practically important normative judgments.”)
VII. CONCLUSION

When it comes to inferences, judges – like people everywhere – should approach the task with great humility.288 Despite best intentions, subconscious mental shortcuts are working away, providing the human mind with information and intuitions that are skewed by each person’s individual experiences. Our very senses can deceive us, influenced by our social and cultural commitments and helped along even by more deliberative thought. And while the cases discussed in this article, and many of the studies of judicial behavior, have focused on difficult and important issues such as race, gender, and politics, there is no reason to think that cognitive shortcuts do not affect all types of decisions in all types of cases.

Yet fact-finding, including the drawing of inferences, is an integral part of the judicial system. It is not confined to juries, or even trials, but is instead a pervasive part of the pretrial period. These important decisions set the parameters for litigation success and failure. (Predictions about particular courts’ heuristic leanings no doubt form part of the basis for forum shopping by both plaintiffs and defendants.) The more we learn from psychologists about the operation of our minds, the less tenable it becomes to stick with the fiction that inferences are purely a matter of neutral logic. The real nature of inference-drawing needs to be confronted.

The Federal Judicial Center is to be commended for the education efforts it has already undertaken, and for including social cognition as a judicial competency. Yet its programs require systemic partners. Knowledge of the issue and tools that encourage reflection must be coupled with efforts to motivate judges to avoid relying too heavily on their individual essential to legal infrastructure in the United States); Arthur R. Miller, From Conley to Twombly to Iqbal: A Double Play on the Federal Rules of Civil Procedure, 60 DUKE L.J. 1, 71–72 (2010) (discussing the importance of private parties’ assisting government agencies in the enforcement of substantive law); Richard B. Stewart & Cass R. Sunstein, Public Programs and Private Rights, 95 HARV. L. REV. 1193, 1289–94 (1982)(discussing the necessity of private rights of action for the enforcement of substantive law).

288 By using this term, the article does not mean to pull in the full body of academic research and theorizing regarding intellectual humility. It is noteworthy, however, that this work is consistent with much of the scholarship cited regarding cognitive and social psychology and with theories of cultural cognition. See generally Mark R. Leary, The Psychology of Intellectual Humility, https://www.templeton.org/wp-content/uploads/2018/11/Intellectual-Humility-Leary-FullLength-Final.pdf (2018)(defining intellectual humility as “recognizing that a particular personal belief may be fallible, accompanied by an appropriate attentiveness to limitations in the evidentiary basis of that belief and to one’s own limitations in obtaining and evaluating relevant information”).
experiences and with the creation of structures that require effortful deliberation. Finally, changes in the way the procedural rules are interpreted and applied could improve the quality of judicial inferences, preserve the power of juries to choose among reasonable inferences, and enhance the depth of the factual record on which those juries will make their decisions. There is no magic bullet to avoid the challenge of humans using human cognition, but we can avoid exalting the “experience and common sense” of individual judges.